PRECISION AGRICULTURE FOR GRAZING AND ANIMAL HEALTH MANAGEMENT: A CASE STUDY IN COLOMBIA

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

in

Engineering

Rodrigo GARCÍA

Universidad EAFIT

Escuela de Ingeniería

Medellín, Colombia

07/09/2023

Director of dissertation research:

Prof. Jose Aguilar	Area of Computing and Analytics

Examining committee:

Prof. María Dolores Rodríguez Moreno	Universidad de Alcalá, Madrid, España
Prof. Juan Carlos Carrascal Velásquez	Universidad de Cordoba, Montería, Colombia
Prof. Luca Carlo Cernuzzi	Universidad Católica Nuestra Señora de la Asunción
	Asunción, Paraguay

Dissertation committee chairperson:

Prof. Jorge Prieto Salazar	Director of the Doctorate in Engineering

President of the Board of the Doctoral Program in Engineering:

Prof. Ricardo Taborda Rios	Dean School of Engineering

Dissertation date : 2023

Abstract

In this research, we address the problem of fattening management and animal health in rotational grazing. For the study of this problem, we have positioned ourselves in the framework of the paradigm of *precision farming*, a technological approach that uses advanced information and communication tools and techniques to optimize agricultural and livestock production processes. In this context, *precision livestock farming* focuses on the use of technologies to improve grazing management and animal health on cattle farms. Some objectives of precision livestock farming are to increase farm efficiency and productivity, improve product quality and reduce production costs. In addition, it also contributes to environmental sustainability by enabling more efficient management of natural resources and reducing negative impacts on the environment. Although precision livestock farming offers many opportunities to improve efficiency and sustainability in livestock production, it also presents challenges that must be addressed for successful implementation. To address this problem, our objective was to develop methodologies, models, and approaches to support decision-making related to productivity management and animal health. To achieve this objective, several sub-objectives were raised, the first one was to develop a precision livestock farming architecture based on emerging technologies (Industry 4.0, artificial intelligence, etc.), the second on developing generic knowledge models of precision livestock farming for animal health and herding management and finally, in the third to develop meta-intelligent models for precision livestock farming in the context of autonomous grazing and animal health management. In general, several research articles were developed to meet the objectives proposed in this thesis. Initially, a review article on the latest trends in precision livestock farming using machine learning techniques was carried out. On the other hand, for the first specific objective, an article was conducted where three autonomous cycles of data analysis tasks based on autonomous computing were proposed for a beef production process for precision livestock farming. To meet the second specific objective, three articles were proposed. The first is a beef cattle weight identification model using machine learning techniques for anomaly detection, the second presented a system for monitoring the cattle fattening process in rotational grazing using fuzzy classification, in the third, a multi-objective optimization model was developed to maximize weight gain of cattle in rotational grazing. Regarding the third objective, three articles were developed, the first one proposed an autonomous cycle of data analysis tasks for the self-supervision of animal fattening in the context of precision livestock farming, and the second article presents a management system for the cattle fattening process in rotational grazing by means of diagnostic and recommendation systems. Finally, the last article proposed the use of the meta-learning paradigm in a cattle weight identification system for anomaly detection. In each article, we evaluated the strategies/models using various datasets. The results showed the capacity of the developed methodologies and models for decision-making in the management of the livestock production process. Specifically, our proposals allow the management of fattening and animal health in rotational grazing, considering, among other things, monitoring, diagnosis, and optimization of the productive process, with good results in performance metrics.

Keywords: Artificial intelligence, Machine learning, Meta-learning, Precision livestock farming, Production management support system, Rotational grazing.

Resumen

En esta investigación, abordamos el problema del manejo del engorde y salud animal en pastoreo rotacional. Para el estudio de este problema, nos hemos posicionado en el marco del paradigma de la *agropecuaria de precisión*, un enfoque tecnológico que utiliza herramientas y técnicas avanzadas de información y comunicación para optimizar los procesos de producción agrícola y ganadera. En este contexto, la *ganadería de precisión* se enfoca en el uso de tecnologías para mejorar la gestión de pastoreo y la salud de los animales en las granjas de ganado bovino. Algunos objetivos de la ganadería de precisión son aumentar la eficiencia y productividad de las granjas, mejorar la calidad de los productos y reducir los costos de producción. Además, también contribuye a la sostenibilidad ambiental, al permitir una gestión más eficiente de los recursos naturales y reducir los impactos negativos en el medioambiente. Aunque la ganadería de precisión ofrece muchas oportunidades para mejorar la eficiencia y la sostenibilidad en la producción de ganado, también presenta desafíos que deben ser abordados para lograr una implementación exitosa. Para abordar este problema, nuestro objetivo fue desarrollar metodologías, modelos y enfoques para apoyar la toma de decisiones en relación con el manejo de la productividad y la salud animal. Para alcanzar este objetivo, se plantearon varios subobjetivos, el primero consistío en desarrollar una arquitectura de ganadería de precisión utilizando tecnologías emergentes (Industria 4.0, inteligencia artificial, etc.), el segundo fue desarrollar modelos de conocimiento genéricos de ganadería de precisión para la gestión de la sanidad animal y el pastoreo, y finalmente, en el tercero desarrollar modelos meta-inteligentes para la ganadería de precisión en el contexto del pastoreo autónomo y la gestión de la sanidad animal. En general, se desarrollaron varios artículos de investigación para cumplir los objetivos propuestos en esta tesis. Inicialmente, se realizó un artículo de revisión de las últimas tendencias de la ganadería de precisión usando técnicas de aprendizaje automático. Por otro lado, para el primer objetivo específico se realizó un artículo donde se propusieron tres ciclos autónomos de tareas de análisis de datos basados en computación autónoma para un proceso de producción de carne de vacuno para la ganadería de precisión. Para cumplir el segundo objetivo específico se propusieron tres artículos. El primero es un modelo de identificación del peso del ganado vacuno mediante técnicas de aprendizaje automático para la detección de anomalías, el segundo presentó un sistema de supervisión del proceso de engorde de ganado en pastoreo rotativo mediante clasificación difusa, y en el tercero se desarrolló un modelo de optimización multi-objetivo para maximizar la ganancia de peso del ganado en pastoreo rotativo. Con respecto al tercer objetivo se desarrollaron tres artículos, el primero propuso un ciclo autónomo de tareas de análisis de datos para la auto-supervisión del engorde de animales en el contexto de la ganadería de precisión, y el segundo artículo presenta un sistema de gestión del proceso de engorde de bovinos en pastoreo rotativo mediante sistemas de diagnóstico y recomendación. Finalmente, el último artículo propuso la utilización del paradigma de meta-aprendizaje en un sistema de identificación del peso del ganado para la detección de anomalías. En cada artículo evaluamos las estrategias/modelos propuestos utilizando diversos conjuntos de datos. Los resultados mostraron la capacidad de las metodologías y modelos desarrollados para la toma de decisión en la gestión del proceso productivo ganadero. Específicamente, nuestras propuestas permiten realizar un manejo del engorde y salud animal en pastoreo rotacional, considerando, entre otras cosas, la supervisión, el diagnóstico, y la optimización del proceso productivo, con muy buenos resultados en las métricas de rendimiento.

Palabras Clave: Inteligencia artificial, Aprendizaje automático, Meta-aprendizaje, Ganadería de Precisión, Sistema de apoyo a la gestión productiva, Pastoreo Rotativo.

Scientific contributions

Several scientific articles were generated and published during the development process of this research project.

Published articles:

- R. García, J. Aguilar, M. Toro, A. Pinto, and P. Rodríguez, "A systematic literature review on the use of machine learning in precision livestock farming" *Computers and Electronics in Agriculture*, vol. 179, p. 105826, 2020. doi.org/10.1016/j.compag.2020.105826
  Q1 Scientific Journal Rankings

- R. García, J. Aguilar, M. Toro, N. Pérez, A. Pinto, and P. Rodríguez, "Autonomic computing in a beef-production process for precision livestock farming", *Journal of Industrial Information Integration*, vol. 31, p. 100425, 2023. doi.org/10.1016/j.jii.2022.100425
  Q1 Scientific Journal Rankings

- R. García, J. Aguilar, M. Toro, and M. Jiménez, "Weight-identification model of cattle using machine-learning techniques for anomaly detection," in *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 01–07, 2021. doi:10.1109/SSCI50451.2021.9659840
  In IEEE Xplore

- C. Benitez, R. García, J. Aguilar, M. Jiménez, and H. Robles, "Supervision system of the fattening process of cattle in rotational grazing using fuzzy classification," in *2022 XVLIII Latin American Computer Conference (CLEI)*, pp. 1–10, 2022. doi:10.1109/CLEI56649.2022.9959950
  In IEEE Xplore

Articles submitted to journals:

- R. García, J. Aguilar, "Using meta-learning in a cattle weight identification system for anomaly detection.", preprint submitted to *Computers and Electronics in Agriculture*, 2023.
  Q1 Scientific Journal Rankings

- R. García, J. Aguilar, and A. Pinto, "An autonomous System for the self-supervision of animal fattening in the context of precision livestock farming", preprint submitted to *Future Generation Computer Systems*, 2023.
  Q1 Scientific Journal Rankings

- R. García, J. Aguilar, and M. Jiménez, "A multi-objective optimization model to maximize cattle weight-gain in rotational grazing", preprint submitted to *Annals of Operations Research*,

2022.
Q1 Scientific Journal Rankings

- R. García, J. Aguilar, and C. Benitez, “Management System for the Fattening Process of Bovines in Rotational Grazing using Diagnosis and Recommendation Systems”, preprint submitted to *CLEI Electronic Journal*, 2023.
  Scopus

Project context

An internal doctoral grant from Universidad EAFIT has fully funded this study. I am grateful to the Vice-Rectory of Discovery and Creation of Universidad EAFIT for their support in this research. The author thanks the farmers for using their knowledge and experience to build the models and interpret the final results.

Acknowledgements

This thesis is the result of a process that began several years ago. This PhD research would not have been possible without the support, collaboration, and contribution of many wonderful people that I am grateful to have in my life.

First, I would like to express my deepest gratitude to Almighty God for being my guide and strength throughout this process of researching and writing this thesis. Thank you for your unconditional love and for giving me the wisdom and knowledge necessary to complete this project. I also want to thank you for the health and blessings that have allowed me to continue moving forward, even in the most difficult moments. Thank you for your constant presence in my life and for being my light in the darkness. To you, God, goes my most sincere gratitude.

I would to thank my parents, Rodrigo Humberto García Barrera and Ruth Claris Hoyos Fajardo, who were my vital support throughout this process. They are the most important people in my life, and it makes me happy that they are proud of me.

I would like to thank my supervisor, Jose Aguilar, Without his help, it would not have been possible to complete this project successfully. From the beginning, he was an invaluable guide, providing me his time and expertise to help me develop my ideas and focus my research. His knowledge of the field and his enthusiasm for teaching have been a great inspiration to me. I especially appreciate his patience and dedication in reviewing and commenting on my work. His suggestions and corrections have significantly improved the quality of my thesis and allowed me to grow academically. I am deeply grateful for his support and confidence in me throughout this process. I feel fortunate to have had the opportunity to work under his guidance.

Finally, I would like to thank Universidad EAFIT for their financial support during my doctoral studies.

Contents

Chapter 1

Introduction and research context

1.1 Problem statement and motivation

In traditional livestock farming, animal species are raised and products are obtained for human consumption; for example, meat and milk. In total, livestock farming provides 33% of the protein consumed in the human diet [1]. Traditional livestock farming faces several challenges today, including climate change, production costs, demand for healthier products, and ever-changing regulations and standards. To meet these challenges, traditional livestock farming must adapt to changes in the market and the environment, seeking new, more efficient and sustainable forms of production.

Specifically, farmers have to manage the comfort conditions in which the herd (group of animals) is kept. In this sense, climate can influence livestock production in two interrelated ways; the first, in a direct way, called comfort, where the animal achieves the balance of physiological processes with the environment, which favors the utilization of feed; and the second, of indirect action, related to the production and supply of feed, which contributes to the degree of comfort [2]. In addition, an important variable to consider is the animal carrying capacity, i.e., the number of animals that can graze in a paddock without affecting forage productivity. Optimum carrying capacity is defined as that at which production per animal per hectare is maximized [3]. Animal carrying capacity and optimum carrying capacity are some of the best indicators of a livestock farm's production [4].

Precisely, the success of any livestock farm depends on having a high optimal stocking capacity, which is achieved with an adequate improvement of paddocks, a correct establishment and maintenance of pastures, an effective paddock rotation, and a balanced management of the animal inventory within the farm. In addition to this, it is necessary to consider variables such as soil quality, rainfall and luminosity, which also affect the animal carrying capacity in the farms [5].

Precision Livestock Farming (PLF) offers several opportunities in this context, including the ability to more effectively monitor and control animal welfare, feed efficiency and early detection of disease [5].

This can result in increased productivity and profitability for producers, as well as improved animal care and product quality. In addition, the technology used in PLF can help reduce the environmental impact of livestock production by minimizing resource use and reducing greenhouse gas emissions.

In this work, five particular challenges were identified: i) determination of the optimal stocking capacity, ii) improvement of paddocks with correct pasture establishment and maintenance, iii) effective herd rotation in paddocks, iv) balanced management of animal inventory within the farm and v) consideration of variables such as soil quality that affect animal carrying capacity on farms. These challenges were addressed under the development of a smart management system based on PLF for beef cattle production that considers pasture and animal characteristics.

The implementation of a smart management system based on PLF offers an efficient and sustainable solution to the challenges mentioned above. This system integrates technologies such as sensors, cameras, and remote monitoring systems to collect data on animal welfare, pasture quality and animal carrying capacity. This data is analyzed to make informed decisions and optimize herd management. In addition, it can detect animal health problems early, which can prevent the spread of disease and reduce the need for costly treatments. Thus, our management system represents a solution to the challenges facing traditional livestock farming today.

1.2 Research objectives

1.2.1 General objective

Build a model for the management of livestock production, considering aspects related to animal health and grazing.

1.2.2 Specific objectives

1. Develop a precision-farming architecture based on the integration of precision agriculture and livestock using emergent technologies (Industry 4.0).
2. Develop generic knowledge models of precision livestock farming for animal health and grazing management.
3. Develop a meta-intelligent model for precision farming in the context of autonomous grazing and animal health management.

1.3 Contributions and research scope

This research focuses on the development of models, methodologies, and computational approaches to support decision-making in rotational grazing management. Several contributions were made in this context. The first contribution was a systematic literature review (SLR) to identify challenges and opportunities for precision livestock research [5]. In this review, we focused on two areas of interest: herding and animal health. For each area, we identify three main challenges for precision livestock development in the context of machine learning, sensors and data sources, and data management. For machine learning, the main challenge is to integrate herding and animal health in approaches that allow the development of diagnostic and prescription models for the prevention and control of animal diseases. Moreover, to explore the utilization of autonomous cycles of data analysis tasks. In terms of sensors and data sources, more affordable sensor technologies are needed to measure humidity, temperature, wind, rain, soil quality and animal health status. In terms of data management, the main challenge is the automatic digitization of all useful farm data, followed by measures to reduce energy consumption and improve data presentation in a user-friendly interface.

Based on the SLR, the need to develop a precision agriculture architecture that integrates precision agriculture and livestock farming using emerging technologies was found to be crucial to meet the growing demand for food production while reducing the environmental impact of agricultural practices. The integration of precision livestock and agriculture can also lead to better monitoring and management of animal health and, ultimately, better quality animal products. The development of such an architecture will be key to achieving sustainable and efficient food production in the face of global challenges such as climate change and population growth. Proposed self-management architecture in beef production is a step towards sustainable and efficient livestock production, aligned with the challenges of climate change and demographic growth. Precise technology in agriculture and livestock farming not only increases productivity, but also contributes to stronger food security and environmental protection in a constantly evolving global context. To address this problem, an architecture for the self-management of beef production farms was proposed as a step toward sustainable and efficient livestock production [6]. This architecture is based on autonomous cycles of data analysis tasks, which provide real-time monitoring and management of the beef production process, allowing for the detection of anomalies and efficient decision-making.

In addition, the development of generic knowledge models for animal health and herd management is essential to advance the field and enable wider adoption of PLF technologies. These models can serve as the basis for the development of more specific and tailored applications. These models can provide valuable information on animal health and grazing patterns, leading to improved management practices and more efficient use of resources. To abort this problem, we developed a cattle fattening process identification system that can be used to detect anomalies in cattle weight gain over time [7]. The identification system was tested using records of animals raised and fattened at a farm located in Montería, Colombia. Four machine-learning techniques were compared for their performance in identifying the ideal weight from real data, including Decision Tree (DT), Gradient Boosting (GB), K-Nearest Neighbors (KNN) regression, and Random Forest (RF). This system can be a useful tool for livestock entrepreneurs to improve the management of their production and increase their profitability. In addition, a diagnostic system was developed for cattle breeders to diagnose weight loss or gain in cattle under a rotational grazing scheme [8]. The system is based on a fuzzy classifier that uses fuzzy reasoning to determine the current situation of the animal given an input, considering both the health status of the animal and the pasture, and genetic algorithms to optimize the rules that characterize the diagnostic process. The proposed system was tested with experimental cases, with promising results. On the other hand, a multi-objective rotational pasture assignment model was also built to improve cattle feeding, considering the best quality forage and minimizing the distance the animals have to travel from one paddock to another. The model considers the amount of forage available and cattle demand, to dynamically allocate animals to paddocks. The effectiveness of the model was evaluated using a discrete simulation of a one-year cattle rotation, and the results indicated that the average weight gain of cattle was higher with the proposed model compared to the traditional rotation method.

Finally, the development of meta-smart models for PLF in the context of autonomous grazing and animal health management is a crucial step toward achieving sustainable and efficient livestock production. These models can be used to interconnect different datasets from multiple farms, enable automatic digitization of farm data, and develop diagnostic and prescriptive models for diseases. In addition, it can enable autonomous cycles of data analysis tasks to support decision-making for grazing and animal health management. To address this problem, several contributions were carried out. Firstly, an architecture for the self-management of beef production farms was proposed [6]. From this architecture, a first autonomous cycle of data analysis tasks was implemented for the self-supervision of animal fattening. This autonomous cycle provides real-time monitoring and management of the animal fattening process, allowing for the detection of anomalies. Moreover, a self-management system for the cattle fattening process in rotational grazing based on diagnostic and recommendation systems has been proposed. Finally, a meta-learning architecture capable of learning, associating, evaluating, and adapting to new data was developed. The use of meta-learning improved the performance of the anomaly detection process by continuously updating the knowledge base with each successful adaptation.

All the contributions made in this research are represented in several research articles. A total of eight (8) scientific articles were generated, of which four (4) are published and the other four (4) are under review.

1.4 Thesis organization

This thesis is presented as a collection of articles developed to meet each of the proposed objectives. In Chapter 2, we conducted an SLR on PLF, which allowed us to collect and evaluate relevant existing information in this field. It identifies the main trends and approaches, as well as knowledge gaps and challenges that still need to be addressed. Chapter 3, Chapter 4, and Chapter 5 correspond to the fulfillment of the first, second and third objectives, respectively. One article was generated for the first objective, three articles for the second objective, and three articles for the last objective. These articles will be presented in each section.

In summary, in Chapter 2, we show an SLR useful for identifying trends, challenges, and research opportunities in PLF using machine learning. Chapter 3 shows an architecture for PLF with autonomic properties fulfilling the objective 1. Moreover, in Chapter 4, three articles corresponding to the second objective proposed in this thesis are presented. The first article corresponds to the development of a cattle weight identification model using machine learning techniques for anomaly detection. The second article corresponds to the development of a system for monitoring the beef cattle fattening process in rotational grazing utilizing fuzzy classification. The third article corresponds to a multi-objective optimization model to maximize the weight gain of cattle in rotational grazing. For Chapter 5, we present three articles corresponding to the third objective proposed in this thesis. In the first article, an autonomous self-monitoring system for animal fattening in the context of PLF was developed. The second article developed a management system for the cattle fattening process in rotational grazing using diagnostic and recommendation systems. The last article presents the use of meta-learning techniques for a livestock weight identification system for anomaly detection. Through meta-learning, the system can adapt and learn from data from multiple farms and environments, improving its performance and accuracy over time.

Finally, Chapter 6 presents a summary of the conclusions of all the papers presented in the previous sections. Finally, in this section, we show the limitations of our research and possible future work.

Chapter 2

State of the art of machine learning in precision livestock farming

2.1 Motivation

In this chapter, we present an SLR that aims to provide a comprehensive overview of the recent advances and applications of machine learning in PLF, with a focus on grazing and animal health. The literature review highlights the potential of machine learning to transform the livestock sector and improve animal welfare. The current use of sensors, software, and techniques for data analysis is also discussed, along with the increasing openness of data sources.

The findings of the review indicate that the use of machine learning in precision livestock farming is still in its early stages of development and faces several research challenges. These challenges include developing hybrid models for the diagnosis and prescription of animal diseases, integrating grazing and animal health issues, increasing the autonomy of precision livestock farming through autonomous data analysis cycles and meta-learning, and bringing together soil and pasture variables for a better understanding of both animal health and grazing behavior. The whole article is in Appendix A.

2.2 Identification of the article

R. García, J. Aguilar, M. Toro, A. Pinto, and P. Rodríguez, "A systematic literature review on the use of machine learning in precision livestock farming" Computers and Electronics in Agriculture, vol. 179, p. 105826, 2020. Q1 Scientific Journal Rankings

2.3 Abstract

This article presents a systematic literature review of recent works on the use of machine learning (ML) in precision livestock farming (PLF), focusing on two areas of interest: grazing and animal health. This review: (i) highlights opportunities for ML in the livestock sector; (ii) shows the current sensors, software, and techniques for data analysis; (iii) details the increasing openness of data sources. It was found that the use of ML in PLF is in a stage of development and has several research challenges. Examples of such challenges are: (i) to develop hybrid models for diagnosis and prescription as a tool for the prevention and control of animal diseases; (ii) to bring together the grazing and animal health issues; (iii) to give autonomy to PLF using autonomous cycles of data analysis tasks and meta-learning; and (iv) to bring together soil and pasture variables because, for both, animal health and animal grazing, the variables used are only behavioral and environmental.

2.4 Link to the full article

https://doi.org/10.1016/j.compag.2020.105826

Chapter 3

Autonomic architecture for beef-production processes

3.1 Motivation

The goal of an autonomic computing architecture in beef production is to reduce the need for manual intervention, improve the accuracy and speed of decision-making, and ultimately, lead to more efficient and sustainable production processes. In this context, this architecture would provide the ability to self-manage, self-configure, self-optimize, and self-heal in real-time, enabling the system to continuously monitor and adjust various aspects of the production process, such as feed, water, and health management, in response to changing environmental and animal health conditions. Additionally, it would provide real-time monitoring and analysis of data, such as animal weight and growth, to allow farmers to make informed decisions about their production processes and improve the overall welfare of their animals. Thus, in this chapter, we present the first paper for the fulfillment of the first objective. The whole article is in Appendix B.

3.2 Identification of the article

R. García, J. Aguilar, M. Toro, N. Pérez, A. Pinto, and P. Rodríguez, "Autonomic computing in a beef-production process for precision livestock farming," Journal of Industrial Information Integration, vol. 31, p. 100425, 2023.

3.3 Abstract

Precision livestock farming (PLF) offers farmers real-time monitoring and management system. PLF provides a real-time warning when something goes wrong so that the farmer can take immediate action to solve the problem. PLF introduces many new challenges and questions that must be resolved. Some of these challenges are related to the integration of grazing and animal health into the beef-production process. This article introduces an architecture for the self-managing of a beef-production farm. In particular, the architecture includes three autonomous cycles of data analysis tasks (ACODAT) that allow beef producers to have adequate coordination, optimization and planning of the productive process, which are: (i) circuit preparation, (ii) animal purchase, and (iii) animal fattening. This article also instantiates, in a farm, the autonomous animal-fattening cycle, as the first step towards efficient and effective beef-production processes. The main contributions of this architecture are (i) the ability to use everything mining to improve the knowledge of the system and decision-making processes, and (ii) three ACODAT for real-time analysis for sustainable and environmentally-friendly livestock production. The results are encouraging since the ACODAT allows smart management of the beef-production process, naturally introducing artificial-intelligence techniques to develop these tasks. Particularly, modeling using ACODAT allows an adequate description of a precision livestock process. Likewise, the preliminary results of some of the tasks of ACODAT are stimulating because they allow evaluating the feasibility of the proposal. For example, a first task for the identification of cattle fattening has a Mean Absolute Error (MAE) of 5.4 kg, which will be used by ACODAT to identify anomalies in the fattening process. The instantiation of the animal-fattening cycle shows the viability and robustness of this proposal.

3.4 Link to the full article

https://doi.org/10.1016/j.jii.2022.100425

Chapter 4

Knowledge models for precision livestock farming

4.1 Motivation

PLF involves the use of technology to monitor and manage the health, nutrition, and behavior of cattle, among other things. The use of knowledge models in this process can help farmers to make data-driven decisions that lead to improved productivity, profitability, and animal welfare. Knowledge models can be used to analyze large amounts of data, identify patterns, and make predictions, among other things, which can then be used to optimize feeding and grazing strategies, disease control, and waste management. With the increasing demand for high-quality and sustainable food production, building knowledge models for PLF is becoming an important aspect of modern agriculture.

Thus, in this chapter, we present three articles to fulfill the second objective. The first article presents one knowledge model for the weight identification of cattle. The second article presents a second knowledge model for the supervision of the rotational grazing cattle fattening process. Finally, the last article describes an assignment model for maximizing the weight gain of rotationally grazed cattle.

4.2 Weight-identification model of cattle

4.2.1 Motivation

The motivation of the livestock weight identification model using machine learning techniques for anomaly detection is to improve the efficiency and accuracy in identifying animals with out-of-normal weight in the livestock industry. By using machine learning techniques, the model can learn and adapt to different situations and conditions, allowing faster and more accurate detection of cattle weight anomalies. This can help farmers and ranchers identify health or nutritional problems in livestock earlier, which in turn can improve the quality and yield of livestock production. Thus, in this section, we present the first article to fulfill the second objective. The complete article can be found in Appendix C

4.2.2 Identification of the article

R. García, J. Aguilar, M. Toro, and M. Jiménez, "Weight-identification model of cattle using machine-learning techniques for anomaly detection", in 2021 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 01–07, 2021.

4.2.3 Abstracts

Cattle raising is an important economic activity, where livestock entrepreneurs keep track of their production and investment costs, to measure production and business profitability, based on cattle weighing. But it's hard for the farmer to tell if the animals they're weighing have gotten the right weight. This paper proposes a framework to identify the fattening process, which can be used to detect anomalies in cattle weight-gain over time. This framework used records of animals raised and fattened at "El Rosario" farm, located in the municipality of Montería (Córdoba-Colombia), to identify the fattening process. The performance of four machine-learning techniques to identify the ideal weight from real data was compared. The algorithms used were Decision Tree (DT), Gradient Boosting (GB), regression based on K-Nearest Neighbors (KNN), and Random Forest (RF). In addition, an outlier-detection process was performed to identify anomalous weights. In general, the results showed that the DT model was the one with the best performance, with an average Mean Absolute Error (MAE) of 5.4 kg.

4.2.4 Link to the full article

https://doi.org/10.1109/SSCI50451.2021.9659840

4.3 Supervision system of the rotational grazing cattle fattening process

4.3.1 Motivation

The motivation of the supervision of the rotational grazing cattle fattening process using fuzzy classification is to improve efficiency and accuracy in monitoring cattle health and growth during the rotational grazing fattening process. By using fuzzy classifiers techniques, the system can adapt to different environmental and grazing conditions, allowing for more accurate and timely monitoring of cattle. This can help farmers and ranchers detect health problems in livestock earlier, make informed decisions about nutrition and grazing management, and improve the quality and yield of livestock production. Thus, in this section, we present the second article to fulfill the second objective. The complete article can be found in Appendix D.

4.3.2 Identification of the article

C.Benitez, R.García, J.Aguilar, M.Jiménez, and H.Robles, "Supervision system of the fattening process of cattle in rotational grazing using fuzzy classification", in 2022 XVLIII Latin American Computer Conference (CLEI), pp. 1–10, 2022.

4.3.3 Abstracts

Cattle breeding has been one of the most important industrial sectors in the world, since it is related to food security and the survival of the human race. Cattle diagnostics is a fundamental procedure for cattle breeders because it allows them to make strategic decisions, such as timely treatment in case of any abnormality (e.g., weight gain in herds, in their paddocks). This article aims to present a system to diagnose weight loss or gain in cattle under a rotational grazing scheme, considering the health status of the animal and the pasture. The diagnostic system is based on a fuzzy classifier that uses fuzzy logic to define the rules that characterize the diagnostic process, and fuzzy reasoning to determine the current situation given an input. In addition, the fuzzy classifier optimizes the rules using genetic algorithms, which modify the membership functions, providing a more accurate system for diagnosis. We tested our proposal with experimental cases, with promising results. The accuracy metrics have high values, indicating a low error rate in terms of false positives. In general, the values of the quality metrics are good, with an accuracy close to 100% and an Area Under the Curve close to 1.

4.3.4 Link to the full article

https://doi.org/10.1109/CLEI56649.2022.9959950

4.4 Assignment model for maximizing weight gain of rotationally grazed cattle

4.4.1 Motivation

The motivation of the multi-objective optimization model for maximizing the weight gain of rotationally grazed cattle is to improve the efficiency and profitability of livestock production. By using optimization techniques, the model can find the optimal combination of factors such as livestock density, grazing duration, grass quality, nutrition, and other relevant factors to maximize the weight gain of rotationally grazed cattle. This can help farmers and ranchers maximize meat production and reduce costs. In addition, the optimization model can also be useful in improving the sustainability of livestock production by optimizing the use of available resources and reducing the negative environmental impacts associated with intensive livestock production. Thus, in this section, we present the third article to fulfill the second objective. The complete article can be found in Appendix D.

4.4.2 Identification of the article

R. García, J. Aguilar, and M. Jiménez, "A multi-objective optimization model to maximize cattle weight-gain in rotational grazing", preprint submitted to *Annals of Operations Research*, 2022.

4.4.3 Abstracts

Rotational grazing can improve cattle feeding considering a series of aspects, such as: (1) the maximum utilization of each hectare of pasture on which the cattle are fed and; (2) the pasture analysis to guarantee the type and the size of the pasture that will serve as feed, among others. The above aspects allow a better-fed cattle, with better weight and meat quality. To implement rotational grazing, it is necessary to carry out forage-utilization practices with criteria associated with the morphophysiology and phenology of the forage species. Many of these data, in the real context of a beef farm, are not used, and most of the decisions made by the farmer are based on experience (successes and failures in productivity). In this proposal, we establish a multi-objective rotational-grazing assignment model based on (1) the best quality forage, and (2) the distance an animal must travel from one paddock to another. At each stage, we estimate the amount of forage as well as the total weight of each batch of animals. Based on pasture yield and cattle forage demand, we propose a dynamic assignment model. To validate the effectiveness of the assignment model, we carried out a discrete simulation of a one-year cattle rotation. Results show that the assignment model generates a statistically higher average weight gain than the one generated by the traditional rotation method.

4.4.4 Link to the full article

See Appendix D.

Chapter 5

Meta-intelligent models for precision farming

5.1 Motivation

The need to develop meta-intelligent models in all areas of artificial intelligence has become a great challenge, either due to the need to make systems more adaptable, or to give them greater autonomy in the processes where they intervene. This does not escape the area of PLF, in such a way as to give a greater degree of intelligence to the processes managed in this context. In this regard, in this chapter, the concept of autonomous cycles of data analysis was used to confer autonomy to various livestock management processes, which are presented in the first section of this chapter. Subsequently, a meta-learning model was proposed to adapt one of the knowledge models presented in the previous chapter, the weigh-identification model, to different data sources.

5.2 Autonomous systems in the context of precision farming.

5.2.1 Autonomous system for the self-supervision of animal fattening

5.2.1.1 Motivation

Autonomous systems for self-monitoring animal fattening aim to improve efficiency and accuracy in the monitoring and management of livestock fattening. By utilizing automation and monitoring technologies, the system can provide constant and accurate monitoring of livestock weight, nutrition, and health without the need for constant human intervention. This can help farmers and ranchers detect health problems in cattle earlier, optimize nutrition and fattening management, and improve the quality and yield of livestock production. In addition, the autonomous system can also be useful in reducing the costs associated with manual livestock monitoring and improving the sustainability of livestock production by optimizing the use of available resources and reducing the negative environmental impacts associated with intensive livestock production. Thus, in this section, we present the first article to fulfill the third objective. The complete article can be found in Appendix F.

5.2.1.2 Identification of the article

R. García, J. Aguilar, and A. Pinto, "An autonomous System for the self-supervision of animal fattening in the context of precision livestock farming", preprint submitted to *Future Generation Computer Systems*, 2023.

5.2.1.3 Abstract

Beef production needs certain levels of autonomy to ensure that animal fattening processes achieve certain sustainability objectives (e.g., financial and environmental). For example, it is required oversight in the animal fattening process, so that stakeholders can make better decisions about what is happening in the fattening process. For monitoring the animal fattening process, this paper proposes an autonomous system. In this paper, this autonomous system is designed and implemented using the methodology for the development of data mining applications called MIDANO, and is tested in a cattle farm simulator that has been developed to reproduce the events of the animal fattening production process. This autonomous system for the self-supervision of the animal fattening process is composed of two data analysis tasks, one to detect anomalies in the fattening of cattle, and another to diagnose this anomaly. The results with real data demonstrate the ability of the proposed supervision system to detect and diagnose anomalies in various conditions (normal, animal health problems, and forage problems in the paddock), and the possible causes of abnormal values in the weight variable. The anomaly detection models have a MAE of the order of 5 kilograms, and the diagnostic model has 95% of Accuracy and 1 of AUC. The results of the experiments are encouraging, as they show that the autonomous system is capable of detecting anomalies and diagnosing them in different operating scenarios. Our system allows giving self-supervision characteristics to a production process.

5.2.1.4 Link to the full article

See Appendix F.

5.2.2 Self-management system of the cattle fattening process

5.2.2.1 Motivation

Improving the efficiency of the cattle fattening process is presented through the development of a management system that takes into account both animal health and pasture quality in the field. This is critical for cattle breeders, as it allows them to make strategic decisions and timely treatments in case of any abnormalities, which in turn improves growth and meat quality while minimizing environmental impact. The proposed system is based on artificial intelligence techniques, such as fuzzy systems and genetic algorithms, for diagnosis, and for pasture recommendation, a classification model, which allows for greater accuracy and efficiency in decision-making. Thus, this section presents the second article to fulfill the third objective. The complete article can be found in Appendix G.

5.2.2.2 Identification of the article

R. García, J. Aguilar, and C. Benitez, "Management System for the Fattening Process of Bovines in Rotational Grazing using Diagnosis and Recommendation Systems", preprint submitted to *CLEI Electronic Journal*, 2023.

5.2.2.3 Abstract

Cattle breeding has been one of the most important industrial sectors in the world, since it is related to food security and the survival of humanity. Management of the cattle fattening process is a fundamental procedure for cattle breeders because it allows them to make strategic decisions, such as timely treatment in case of any abnormality (e.g., weight gain in herds, in their paddocks). This article aims to present a management system for the cattle fattening process under a rotational grazing scheme, considering the health status of the animal and the pasture, which should diagnose weight loss or gain in bovines and recommend actions when is required. The diagnostic process is based on a fuzzy system that defines rules that characterize the diagnostic process to determine the current situation given an input. Furthermore, the fuzzy classifier optimizes its rules by means of genetic algorithms by modifying its membership functions, providing a more accurate system for diagnosis. On the other hand, the recommendation system is based on a classification model of pasture crops, in which the best pasture is recommended given the soil variables. We tested our proposal with experimental cases, with promising results. For the fuzzy classifier, the accuracy metrics are good, with values of accuracy close to 100% and of Area Under the Curve close to 1. For the classification model were used several machine learning techniques, resulting in the best classifier the random forest technique, with an accuracy of 98.61%.

5.2.2.4 Link to the full article

See Appendix G.

5.3 Meta-Learning in the context of PLF

5.3.1 Motivation

The motivation for using Meta-Learning in the context of PLF is to continuously improve the accuracy and performance of the decision-making system over time. This can be achieved by updating the knowledge base of the system with new data and experiences, or by allowing the system to adapt and learn from past successes and failures. In PLF, by using Meta-Learning, a system can continuously improve its accuracy and performance to identify patterns and anomalies in the fattening process based on previous observations. Specifically, it can learn from previous experiences (for example, learned on other farms) and unlearn from outdated models to update them. This section presents the third article to fulfill the third objective. The complete article can be found in Appendix H.

5.3.2 Identification of the article

R. García, J. Aguilar, "Using meta-learning in a cattle weight identification system for anomaly detection.", preprint submitted to *Computers and Electronics in Agriculture*, 2023.

5.3.3 Abstract

Weighing management in cattle farming is important for farmers, as it allows them to accurately monitor the growth and development of their animals. It is also a valuable tool that allows farmers to maximize production and the welfare of their animals. However, it is difficult for the farmer to detect if the herd of animals being weighed is gaining the ideal weight for a given breed and age. In addition, normally, when a new breed of cattle is introduced to a farm, there is very little data. This article proposes a meta-learning framework (MTL) for identification models used in the fattening process of animals to detect anomalies in cattle weight. The proposed MTL framework has a knowledge base of Meta-Models on Identification models based on machine learning techniques, which is used to select the identification model to use when a new breed of cattle arrives on the farm. This knowledge base is updated, either because a previous identification model has been successfully adapted to the new breed, or a new identification model has had to be generated, allowing the framework to continuously improve its performance over time. Particularly, this article presents in detail the process of adaptation of the previous identification models to new breeds carried out by our MTL framework. Besides, to test our approach a case study is presented, using records of animals raised and fattened at the "El

Rosario" farm, located in the municipality of Monteria (Córdoba-Colombia). The results are very encouraging in terms of the ability of our framework to adapt the identification models to different possible scenarios in a process of detecting anomalous weights. In general, the identification models generated with our proposal had an R^2 of 90.8%, which suggests that the models can explain the variability observed in the data.

5.3.4 Link to the full article

See Appendix H.

Chapter 6

Conclusions

In this chapter, we present a summary of the results of all the work presented above. In addition, we show limitations and research opportunities for the future.

6.1 Summary

As a first main conclusion, in our systematic literature review, we highlight the current state of the art of machine learning techniques used in precision livestock for herding and animal health monitoring. Most of the works focus on sheep and cattle, using classification techniques to analyze animal behavior and variables such as feeding behavior. However, there are still important limitations in the use of variables, as soil variables and animal production factors are not adequately considered. In addition, most animal health studies have focused on monitoring animal behavior with classical Machine Learning techniques for classification, and there is no work on diagnosis, prediction, or prescription of treatments for diseases. We identify the urgent need to address these limitations that can incorporate different variables and production management to improve the efficiency of precision livestock farming.

To assist farmers in these tasks, we developed a precision farming architecture, which integrates and interoperates different actors in the context of PLF. We include autonomous self-planning cycles that can adapt to new environmental constraints and changes in forage demand. We also highlight the use of mining techniques as a key factor in decision-making processes. Furthermore, we present a case study of rotational grazing as an example of how to incorporate self-management into a beef production process.

In addition, we developed generic knowledge models based on data mining and artificial intelligence for animal health and grazing in the PLF context. Particularly, an identification model for detecting anomalies in beef cattle fattening processes. The model yields promising results in detecting anomalous observations of cattle body weight, which is an effective solution for the beef production industry. Furthermore, the developed diagnostic system, based on fuzzy logic and genetic algorithms, has proven to be an effective tool for diagnosing the weight loss or gain of cattle in rotational grazing regimes. Finally, the rotational pasture allocation model developed is a tool that maximizes the use of each hectare of pasture and considers the distance traveled by each animal. The results indicate that the model generates a higher average weight gain than the traditional rotational method.

Importantly, meta-intelligent precision livestock modeling in the context of autonomous grazing and animal health management is a novel approach to developing adaptive knowledge models for meat production, which allows systems to continuously improve performance over time. For example, our meta-learning model uses a knowledge base of machine learning models to detect anomalies in animal weights, and updates with each successful adaptation to improve performance. This allows a continuous learning process from the data, which translates into greater efficiency and profitability for producers. In addition, we developed an autonomous system for managing the animal fattening process in rotating paddocks, which includes a fuzzy diagnostic system and a recommendation system based on a grass classifier model. The system has the potential to be used in livestock disease diagnosis and paddock performance optimization. We also developed a self-monitoring system to detect and diagnose anomalies in animal weight. The system combines several machine learning techniques to achieve its objectives, being flexible and adaptable. The results indicate that the system is effective in detecting and diagnosing anomalies.

In general, the results of the experiments are encouraging, as they show that our different propositions can plan, monitor, identify, detect and diagnose, among other things, in different operational scenarios, allowing better decisions to be made in meat production.

6.2 Limitations and future work

We developed strategies or approaches to support decision-making in grazing and animal health management. With the results of the present research, we were able to meet the proposed objectives.

However, this research has some limitations, particularly, in the precision agriculture architecture based on the integration of agriculture and livestock using emerging technologies. One of the main limitations of the proposed architecture is the cost associated with the sensors for data acquisition, which can be a major obstacle to its implementation [9, 10]. In addition, ensuring optimal sensor distribution, stable power flow, and internet connectivity in farms, may also pose a challenge [11]. Another potential limitation is the assumption that the ideal grazing plan for a specific breed of cattle is similar in different contexts, which may not always be the case [12].

Furthermore, the developed generic knowledge models for animal health and grazing in the context of PLF are only applicable to rotational grazing, for a cattle population from the tropics, and for two climatic seasons (summer and winter). Therefore, this system can only be used in contexts with the following characteristics: in farms that use rotational grazing, regardless of the size of the cattle herds, and that the climatic seasons are summer and winter (for the tropics) where winter is rainy and summer is dry. It is important to note that this system may not be applicable to farms with different management practices, or in regions with different climatic conditions.

Additionally, the developed meta-intelligent models in the context of PLF for autonomous grazing and animal health management require a large amount of data to train the initial models. Also, it is assumed that ideal weight growth curves for specific breeds of cattle are similar in all contexts. Moreover, our architecture may not be suitable for farms that do not have access to data analysis tools.

Future work on the architecture of precision agriculture based on the integration of agriculture and livestock using emerging technologies could be directed at developing an optimal sensor distribution model, to obtain adequate and sufficient coverage of the area to be monitored. If sensors are not properly distributed, critical data may be lost or the quality of the data collected may be low, which may compromise the effectiveness of the livestock production monitoring and management system.

Moreover, in the development of generic knowledge models for animal health and herding based on fuzzy classifiers, it is possible to extend the process of rule adaptation (currently only based on trapezoidal membership functions). For example, allowing other membership functions (e.g., Gaussian), or even, the possibility of making changes in the variables used in the rule antecedent. Along the same lines, other possible extensions are to use more context variables (e.g., explicitly the weather in the antecedents), or to allow varying the number of fuzzy sets in the fuzzy variables (e.g., more or fewer states to characterize the fattening process).

In the context of our meta-learning architecture, testing different synthetic data generation algorithms for the PLF context and evaluating their performance and behavior in our architecture is a future work. In addition, incorporating prescription models to define how to act in the event that an anomaly is diagnosed in the meat production process is a future extension to our autonomic cycles.

References

[1] K. Suryawanshi, S. Redpath, Y. Bhatnagar, U. Ramakrishnan, V. Chaturvedi, S. Smout, and C. Mishra, "Impact of wild prey availability on livestock predation by snow leopards," *Royal Society Open Science*, vol. 4, no. 6, 2017.

[2] M. Velez-Terranova, R. Molina, H. Sánchez, R. Campos, and S. Perilla, "Influence of climatic conditions on tympanic temperature and milk production in grazing cows," *Journal of Animal Behaviour and Biometeorology*, vol. 9, no. 4, 2021. cited By 0.

[3] S. Landau, L. Dvash, Y. Yehuda, H. Muklada, G. Peleg, Z. Henkin, H. Voet, and E. Ungar, "Impact of animal density on cattle nutrition in dry mediterranean rangelands: A faecal near-ir spectroscopy-aided study," *Animal*, vol. 12, no. 2, pp. 265–274, 2018. cited By 7.

[4] A. Skonhoft, "On the optimal exploitation of terrestrial animal species," *Environmental and Resource Economics*, vol. 13, pp. 45–57, Jan 1999.

[5] R. García, J. Aguilar, M. Toro, A. Pinto, and P. Rodríguez, "A systematic literature review on the use of machine learning in precision livestock farming," *Computers and Electronics in Agriculture*, vol. 179, p. 105826, 2020.

[6] R. García, J. Aguilar, M. Toro, N. Pérez, A. Pinto, and P. Rodríguez, "Autonomic computing in a beef-production process for precision livestock farming," *Journal of Industrial Information Integration*, vol. 31, p. 100425, 2023.

[7] R. García, J. Aguilar, M. Toro, and M. Jiménez, "Weight-identification model of cattle using machine-learning techniques for anomaly detection," in *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 01–07, 2021.

[8] C. Benitez, R. García, J. Aguilar, M. Jiménez, and H. Robles, "Supervision system of the fattening process of cattle in rotational grazing using fuzzy classification," in *2022 XVLIII Latin American Computer Conference (CLEI)*, pp. 1–10, 2022.

[9] T. Mizik, "How can precision farming work on a small scale? a systematic literature review," *Precision Agriculture*, pp. 1–23, 2022.

[10] P. Balogh, A. Bai, I. Czibere, I. Kovách, L. Fodor, Á. Bujdos, D. Sulyok, Z. Gabnai, and Z. Birkner, "Economic and social barriers of precision farming in hungary," *Agronomy*, vol. 11, no. 6, p. 1112, 2021.

[11] M. Zhang, X. Wang, H. Feng, Q. Huang, X. Xiao, and X. Zhang, "Wearable internet of things enabled precision livestock farming in smart farms: A review of technical solutions for precise perception, biocompatibility, and sustainability monitoring," *Journal of Cleaner Production*, vol. 312, p. 127712, 2021.

[12] D. M. Freebairn, A. Ghahramani, J. B. Robinson, and D. J. McClymont, "A tool for monitoring soil water using modelling, on-farm data, and mobile technology," *Environmental modelling & software*, vol. 104, pp. 55–63, 2018.

Appendix A

A systematic literature review on the use of machine learning in precision livestock farming

Computers and Electronics in Agriculture 179 (2020) 105826

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

A systematic literature review on the use of machine learning in precision livestock farming

Rodrigo García[a,c], Jose Aguilar[b,c,*], Mauricio Toro[c], Angel Pinto[a], Paul Rodríguez[d]

[a] *Department on Informatics and Systems, Universidad del Sinú, Montería, Colombia*
[b] *CEMISID, Universidad de Los Andes, Merida, Venezuela*
[c] *GIDITIC, Universidad EAFIT, Medellin, Colombia*
[d] *Faculty of Economics, Administration and Accounting, Universidad del Sinú, Montería, Colombia*

A R T I C L E I N F O

Keywords:
Precision livestock farming
Animal health
Grazing
Artificial intelligence
Machine learning
Big data analytics
Industry 4.0

A B S T R A C T

This article presents a systematic literature review of recent works on the use of machine learning (ML) in precision livestock farming (PLF), focusing on two areas of interest: grazing and animal health. This review: (i) highlights opportunities for ML in the livestock sector; (ii) shows the current sensors, software and techniques for data analysis; (iii) details the increasing openness of data sources. It was found that the use of ML in PLF is in a stage of development and has several research challenges. Examples of such challenges are: (i) to develop hybrid models for diagnosis and prescription as a tool for the prevention and control of animal diseases; (ii) to bring together the grazing and animal health issues; (iii) to give autonomy to PLF using autonomous cycles of data analysis tasks and meta-learning; and (iv) to bring together soil and pasture variables because, for both, animal health and animal grazing, the variables used are only behavioral and environmental.

1. Introduction

In traditional livestock farming, animal species are bred, and products for human consumption are obtained; for instance, meat and milk. Overall, livestock provides 33% percent of the protein consumed in the human diet (Suryawanshi et al., 2017). Recently, it has emerged the concept of *Precision Livestock Farming (PLF)*: A holistic approach that adds *information and communication technologies (ICT)* to improve the farming process. PLF plays an important role in the fourth industrial revolution, also known as Industry 4.0. PLF uses ICT to reduce investment costs and increase both, production and animal health (Banhazi et al., 2012).

In traditional livestock farming, decisions are often based –only– on the experience of the producer. In PLF, such decisions are based on quantitative data, such as liters of milk per milking. In addition, quantitative data can be obtained in real-time. To obtain and study such data, in real-time, PLF systems use data analysis, *machine learning (ML)*, control systems, and ICT (Banhazi et al., 2012).

The main objectives of PLF are: (i) to identify the most appropriate livestock feeding, (ii) reduce environmental impact through efficient resource management, (iii) manage crop processes to make a perfect synergy with livestock feeding, (iv) ensure food safety through traceability (documentary record from production to consumption) of products, and (v) improve animal health and crop efficiency (Pomar et al., 2011).

PLF must improve the efficiency of production systems. To improve efficiency, it is essential to, correctly, manage data generated every day in livestock farms (Suryawanshi et al., 2017). To manage data, such systems must perform collection, processing, analysis and distribution of information. A correct data management can result in improved productivity, in terms of grazing lot management, livestock nutrition, and animal health.

Technology over the years has made easier to carry out traditional farm activities. Specifically, in livestock production, it is now possible to process data collected daily related to animal control (Vranken and Berckmans, 2017). ICT applied to the livestock sector has made possible to exploit this data to predict and describe the behavior associated with a more efficient animal production (Espinosa et al., 2016). To make such predictions, ML is often used. In general, this information about animal behavior allows determining the needs of the animals, providing personalized and optimal attention for the benefit of the production (Banhazi and Black, 2009).

* Corresponding author.
E-mail addresses: rjgarciah@eafit.edu.co (R. García), aguilar@ula.ve (J. Aguilar), mtorobe@eafit.edu.co (M. Toro), angelpinto@unisinu.edu.co (A. Pinto), deccienciaseconomicas@unisinu.edu.co (P. Rodríguez).

https://doi.org/10.1016/j.compag.2020.105826
Received 18 June 2020; Received in revised form 8 September 2020; Accepted 5 October 2020
Available online 20 October 2020
0168-1699/

At present, PLF seeks, through technological solutions in agricultural and livestock production systems, to supply adequate food for the expected world population of more than nine billion inhabitants by 2050 (Perakis et al., 2020). PLF allows the sector to be extended to sustainable livestock production, by considering production and animal health. In addition, PLF allows producers to maintain an optimum number of animals per farm, find prompt solutions to animal diseases, and define a more efficient production model (Berckmans, 2014). In this context, a *systematic literature review (SLR)* is needed to determine the state-of-the-art in PLF related to artificial intelligence, and, in particular, current ML techniques used in PLF, and the challenges for the following years.

To understand ML approaches and challenges, in the following years, in the context of PLF, two aspects are discussed in this paper. Firstly, a review on the state-of-the-art of PLF based on the following topics: works on ML for grazing and for animal health. According to Van der Burg et al. (2019), these two topics represent research challenges in the field of PLF. Secondly, a discussion of current challenges for PLF related to ML.

The main contributions of this article are the following. First, a presentation of the most widely used ML techniques, in the last five years, for the analysis of grazing and animal health. Second, an overview of the most used techniques of PLF data acquisition in the literature. As a final result, an overview of the works related to ML and PLF, and future research directions according to our analysis of the previous researches.

The rest of this article is organized as follows. Section 2 introduces the methodology of the SLR, including an analysis for the need of a new SLR in PLF. Section 3 presents the results of the review. Finally, Section 4 concludes the document by a general discussion around how the research questions have been answered, and presents the current limitations and challenges.

2. Methodology

The methodology used is of SLR, which divides the process into four phases (Torres-Carrion et al., 2018): (i) identification of the need for revision, (ii) definition of a review protocol, (iii) conducting the review, and (iv) development of an analysis of the review.

In the first phase, the goal is to identify why a new SLR is needed in the domain. In the second phase is set up the information collection process. For that, it is specified the questions that guide the research and the search strategies; the criteria for the inclusion or exclusion of documents, the evaluation of the relevance and quality of the documents; and, finally, the queries for academic data sources. In the third phase is carried out an exhaustive review of each document to determine if such documents answer the research questions. Documents that do not meet the expected characteristics are discarded. Finally, in the last phase is presented a general analysis for each research question.

2.1. Identification of the need for revision

In this section, previous SLRs in the domain of PLF are presented. The reviews analyzed in this section are summarized in the following categories: models for farming enterprise, interoperable standards in extensive livestock farming systems, dairy energy, and animal health and animal monitoring.

2.1.1. Models for farming enterprise

O'Grady et al., in O'Grady and O'Hare (2017), presented an overview of models for a farming enterprise. Some examples of the models are *Great Plains Framework for Agricultural Resource Management (GPFARM)*, from North America, applied to the entire farm domain; GRAZPLAN, from Australia, applied to grazing enterprises; EcoMod, from Australia and New Zealand, applied to pasture management; *Agricultural Production Systems Simulator (APSIM)*, from Australia, used for crop modeling; *National Research Council (NRC)*, from North America, focused on nutrition animal factors; *Nordic feed evaluation system (Norfor)*, from Scandinavia, focused on nutrition animal and feed factors; *Total Dry Matter Intake Index (TDMI)*, from Finland, implemented for nutrition dry matter Intake animal and feed factors; Biopara-Milk, from the United Kingdom; and, finally, Karoline, from Scandinavia, used to analyze the impact of feeding on rumen pH in dairy cow for their nutrition, milk production, digestion, and gas emissions. GRAZPLAN, EcoMod and APSIM models are focused on grazing, pasture and crops, respectively. O'Grady et al. expressed that the difficulty of adoption of these models, by individual farmers and agricultural enterprises, depends on the usability and the identification of best practices in their activities.

2.1.2. Interoperable standards in livestock farming systems

Bahlo et al. (2019) reviewed interoperable standards in extensive livestock farming systems, and concluded that there is a need for a new type of decision support tool. Bahlo et al. argued that both, farm-centric and farmer-centric approaches are needed. Particularly, they also concluded that a consensus is needed on data exchange standards to prove the value of shared data at the farm scale (commercial benefit) and a regional scale (public good).

2.1.3. Dairy energy

Shine et al. (2020) focused on summarizing and reviewing dairy-energy research from the monitoring, prediction, and analysis points of view. According to Shine et al., dairy-energy prediction models have been frequently used throughout the literature to conduct dairy- energy analysis, and to estimate the impact of changes in the infrastructural equipment and managerial practices.

2.1.4. Animal health

On the domain of animal health, Mcloughlin et al. (2019) discussed current trends in sound analysis for ecology and conservation. Mcloughlin et al. detailed the vocalizations produced by three of the most important farm livestock species: chickens (*Gallus gallus domesticus*), pigs (*Sus scrofa domesticus*) and cattle (*Bos taurus*). Mcloughlin et al. described methods to monitor animal health based on the sound, with the potential to be automated for large-scale farming. In addition, Benjamin and Yik (2019) conducted a review aimed at veterinarians and pig specialists, describing machine learning algorithms, such as pig-face recognition –using convolutional neural networks. They also identified the most relevant sensors for measuring animal health: cameras (2D and 3D), microphones, thermistors and accelerometers. Finally, they discussed how these technologies can be used to improve pig health.

2.1.5. Animal monitoring

Recent technologies to monitor animals are based on machine-learning and computer-vision algorithms. There are four SLRs focused on animal monitoring. First, Norton et al. (2019) discussed the main technology-oriented approaches to animal monitoring and showed how image and sound analysis can be used to build *digital representations* of animals. Second, Milan et al. (2018) proposed a discussion on how ML algorithms could be used to monitor internal and surface temperatures, breathing rate, sweat rate, gait pattern, behavior, physical dimensions, weight and body-condition score of dairy cows. Third, Astill et al. (2020) discussed the areas of impact that new smart-sensor technologies will have on poultry operations, and described how sensor technology is related to big-data analytics and the *Internet of Things (IoT)*, and how these technologies can enhance the productivity of the poultry industry. Finally, Dominiak and Kristensen (2017) presented methods that classify or prioritize the alerts that occurred in the herds. They suggested that future researches should focus on alternative approaches of detection models, using the prior probability or risk of a condition to occur.

2.1.6. Summary

Table 1 shows the problems addressed by recent literature reviews on PLF. In general, the tasks related to PLF have focused on the

Table 1
Summary of recent SLRs.

Article	Objective
O'Grady and O'Hare (2017)	Successful enterprise farm models.
Bahlo et al. (2019)	Interoperable data standards in farming systems.
Shine et al. (2020)	Dairy energy for monitoring, prediction and analysis.
Mcloughlin et al. (2019)	Sound classification for ecology and conservation.
Benjamin and Yik (2019)	Pig health with sensors and ML.
Norton et al. (2019)	Image and sound to build digital animal-representations.
Milan et al. (2018)	Precision dairy farming.
Astill et al. (2020)	New sensor technologies on poultry operations.
Dominiak and Kristensen (2017)	Methods that prioritize alerts occurred in the herd.

classification of animal behavior and monitoring; nonetheless, previous literature reviews do not analyze the use of ML in grazing and animal health like contexts of application. Our article reviews different ML techniques, on the context of PLF, that have been used for the analysis of grazing and animal health, as well as the different forms of data acquisition for training such ML models. Thus, there are no SLRs about the use of ML models in PLF nor in the context of grazing and animal health.

2.2. Document selection process

In this section, activities carried out for the SLR are introduced. First, research questions and search strategies are presented. Second, document selection, inclusion and exclusion criteria are explained. Finally, quality evaluation and a summary of the selection process are detailed.

The research questions, in the two areas of interest of this research, are: (i) How have systems for grazing been developed based on ML in PLF? and (ii) How has ML been used for animal health in PLF? The keywords for these questions are summarized in Table 2.

The boolean search equations generated for this SLR combine the previous keywords to answer each research question. The sources used were Google Scholar, IEEE Xplore, Scopus and Springer databases. Search equations are summarized in Table 3. The initial documents, obtained with these search equations, were 171.

Table 4 shows the inclusion and exclusion criteria used in this SLR. Each inclusion criterion is assigned a nomenclature CIi.

Table 5 shows the results after applying the inclusion and exclusion criteria to the queries in Google Scholar, IEEE Xplore, Scopus and Springer databases. The total was of 54 articles, for the period from 2015 to 2020.

In what follows, a set of questions is presented to ensure the quality of the documents related to each research question. Such questions allow evaluating the quality of the documents. With respect to ML, for grazing in PLF, there are two questions for quality evaluation: (i) Does the document explain the use of ML in grazing? and (ii) Does the document describe a PLF management system based on ML? With respect to ML for animal health in PLF, the questions are: (i) Does the document describe how ML enriches data collection in animal health in PLF? and (ii) Does the document describe how sensors are used to measure animal health in the context of PLF?.

Fig. 1 shows the selection process to find the relevant works for this research.

Table 2
Terms or keywords for the research equations.

Precision Livestock Farming			
A1	Animal health	**A2**	Precision livestock farming
A3	Animal welfare	**A4**	Grazing
Machine Learning			
B1	Machine learning	**B2**	Big data
B3	Data models		

Table 3
Search equations for each research question.

Question	Search equation	Documents
Grazing	A4 AND A2 AND (B1 OR B2 OR B3)	92
Animal Health	(A1 OR A3) AND A2 AND (B1 OR B2 OR B3)	79
	Total	171

Table 4
Inclusion and Exclusion Criteria.

Inclusion Criteria	CI1 Scientific articles, conference proceedings. CI2 Publications are in English CI3 Publications after the year 2015 CI4 Belong to sub-area of Computer Science CI5 Belong to sub-area of Engineering
Exclusion Criteria	Publications before 2015. Non-digital publications. Publications not available for full review.

Table 5
Summary of equations for the research questions after inclusion and exclusion criteria.

Question	Search equations with inclusion and exclusion criteria	Result
Grazing	(A4 AND A2 AND (B1 OR B2 OR B3)) AND CI1 AND CI2 AND CI3 AND CI4 AND CI5	26
Health	(A1 OR A3) AND A2 AND (B1 OR B2 OR B3)) AND CI1 AND CI2 AND CI3 AND CI4 AND CI5	28
	Total Documents	**54**

For question on ML for grazing, the quality assessment left 15 works; and for question on ML for animal health, the quality assessment left 20. At the end of the selection process, there was a total of 35 documents to be analyzed.

2.3. Preliminary results

The preliminary results are divided into two parts. First, the results regarding ML for grazing in PLF and, after, the results regarding ML for animal health in PLF.

2.3.1. Machine learning for grazing in the PLF context

The purpose of grazing is: (i) to maintain a high production of good quality fodder for the longest period of time, (ii) to maintain a favorable balance between fodder species, and (iii) to obtain an efficient utilization of the fodder produced, achieving a profitable livestock production (Bell et al., 2014).

Regarding the question of how ML has been used to improve grazing in PLF, 15 documents were selected, from which 12 are in journals and 3 are in conferences. Fig. 2 shows Australia with the largest production of documents focused on grazing and, especially, on sheep. The largest production of documents may be the impetus given by the *cooperative research center (CRC) program* that has developed, in recent years, improvements in technologies for identification and electronic registration, allowing farmers to manage their sheeps with greater accuracy. The Australian farmers weigh their animals frequently. Live weight –and its variations in time– give farmers a good idea of the nutritional status of the animals. The use of electronic precision scales, with *radio frequency communication interface (RFID)* and data-storage equipment, transforms a tedious process –with numerous errors– into something simple, fast and natural (Caja et al., 2016).

The United States is the country that produces the most beef in the world: 12,700 metric tons of beef. Such a production is followed by Brazil, with 10,000 tons (Porter, 2019). In most of these countries, beef

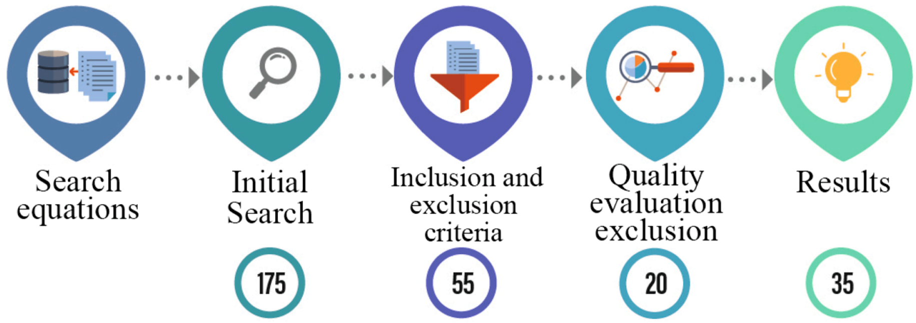


Fig. 1. Selection process of articles for the SLR.

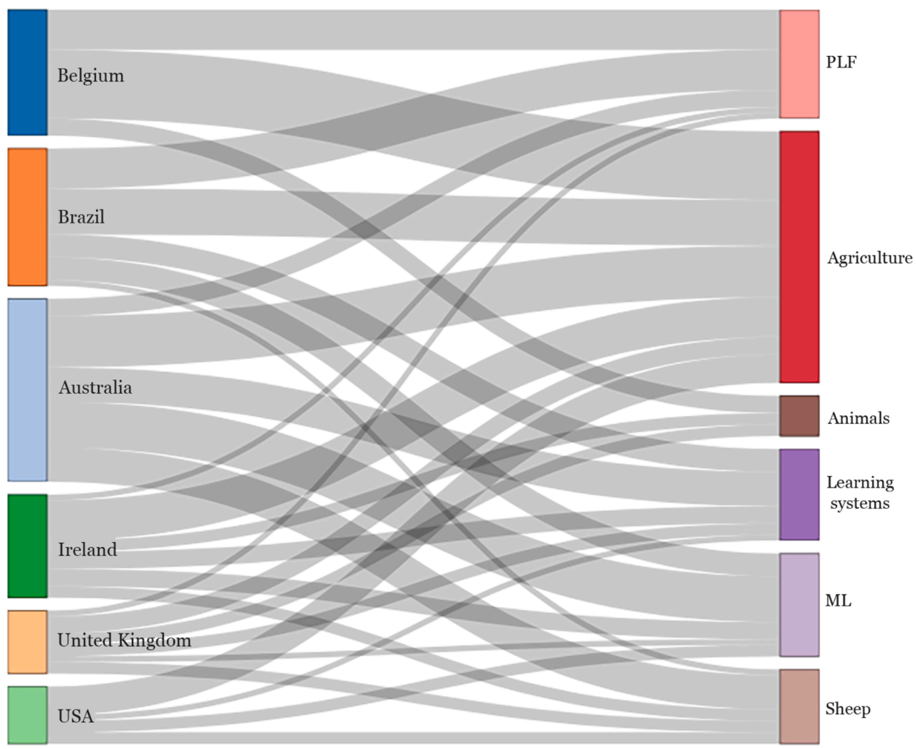


Fig. 2. Countries by keywords for the research question of ML for grazing in PLF.

exports also have a great weight in their economies; for this reason, researches on PLF, ML and pasture agriculture –in these countries– is very important, as shown in Fig. 2.

2.3.2. *Machine learning for animal health in the PLF context*

Animal health is the state of an animal to cope with its environment. Animal health is related to all coping mechanisms, involving the physiology, behavior, feelings and the responses to pathologies (Machado and Silva, 2019). Ensuring a safe, sufficient and nutritious food supply for a population that is increased so rapidly, depends on healthy and productive animals (Overgaauw et al., 2020).

Regarding the question of how ML has been used to improve animal health in the PLF domain, 20 documents were selected, of which 11 are in journals, 8 are in conferences and 1 is a literature review. Fig. 3 shows that there is widespread interest in this issue, demonstrated by the growing public concern for animal health in most countries of the world. Animal health is an important part of well-being (Singer et al., 2019). Not surprisingly, countries such as Brazil, China, and Australia, which have a large production of meat, are interested in measuring behavior and monitoring well-being, using ML techniques. Measuring behavior and monitoring well-being helps them to increase production and export a higher-quality product.

3. Review report

The results are divided into two parts. First, the results regarding ML for grazing, in the PLF context, and, after, the results regarding ML for animal health in the PLF context.

3.1. *Machine learning for grazing in PLF*

In this section, the analysis of the documents is divided into works about the classification of animal behavior and about the animal monitoring.

3.1.1. *Classification of animal behavior*

The following documents focus on the classification of animal behavior. The first category describes the documents that focus on how to classify livestock behavior in general. The second category presents documents –specifically– on the classification of grass intake and rumination activities.

Livestock-behavior classification. Animal behavior can be

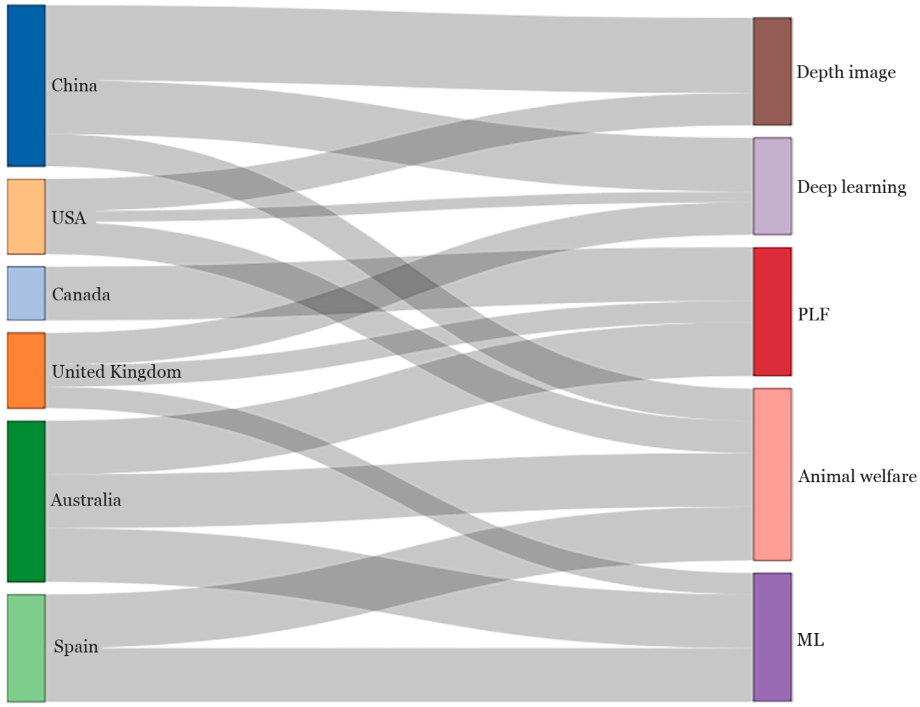


Fig. 3. Countries by keywords for the research question of ML for animal health in PLF.

classified into different categories and with different sensors. Vázquez-Diosdado et al. (2019) presented a combined offline *k-nearest neighbors (KNN)* algorithm and an online learning algorithm, which is deemed by the authors as a useful mechanism for long-term in-the-field monitoring systems. The proposed algorithm classifies relevant sheep behavior using information from an embedded edge device that includes a tri-axial accelerometer and tri-axial gyroscope sensors.

Another alternative to KNN algorithm for behavior classification is a hierarchical ML. Suparwito et al. (2019) proposed the use of a hierarchical ML method to classify livestock behavior. Firstly, they classified livestock behavior into two behavioral categories: active or inactive. Each of the two categories is then broken down at the next level into more specific behavioral categories. Secondly, they tested the proposed methodology using two commonly used classifiers: *Random Forest (RF)*, *Support Vector Machine (SVM)* and, a newer approach, *Deep Belief Networks (DBN)*. Finally, results showed that the hierarchical classification technique works better than the conventional approach.

Further than the classification of behavior in active or inactive, one can classify behavior using an inventory of actions exhibited by animals. Such inventory is known as an *ethogram*. Fogarty et al. (2020) explored feature creation and ML algorithms to provide the most accurate behavioral classification using an ear-borne accelerometer in extensively-grazed sheep. Nineteen derived-movement features, three epochs (5, 10 and 30 s) and the next four ML-algorithms, were assessed: *Classification and Regression Trees (CART)*, SVM, *Linear Discriminant Analysis (LDA)* and *Quadratic Discriminant Analysis (QDA)*. The behavior classification was also evaluated using three different ethograms, including detection of (i) grazing, lying, standing, walking; (ii) active and inactive behavior; and (iii) body posture. The detection of the four mutually-exclusive behaviors (grazing, lying, standing and walking) was the most accurately performed, using a 10 s epoch, by SVM (76.9%). The most accurately was detected, using a 30 s epoch, by CART (98.1%). LDA, using 30 s epoch, was superior for detecting posture (90.6%).

The concept of ethograms used in sheep can also be extended to dairy cows. Vázquez D. et al. (2015) developed a DT algorithm that uses tri-axial accelerometer data from a neck-mounted sensor to classify biologically important behavior in dairy cows and detect transition events between lying and standing. Data was collected from six dairy cows that were monitored, continuously, for 36 h. Direct visual observations of each cow were used to validate the algorithm. The results show that the DT algorithm is able to –accurately– classify three types of biologically relevant behaviors: lying (77.42% sensitivity, 98.63% precision), standing (88.00% sensitivity, 55.00% precision), and feeding (98.78% sensitivity, 93.10% precision). Transitions between standing and lying were also detected –accurately– with an average sensitivity of 96.45% and an average precision of 87.50%.

Another approach to define an ethogram is using unsupervised learning. The objective of Achour et al. (2019) was to develop an effective unsupervised-classification model of data collected by *Inertial Measurement Units (IMU)*, attached to the back of dairy cows housed in free-stall. Data was aggregated according to different sampling frequencies and segmentation windows. The different times of lying, standing, lying down, standing up, walking and stationary behaviors were observed and recorded in real-time. The designed classification model is based on univariate and multivariate *Finite Mixture Models (FMM)* and DT. The valid transitions between standing and lying behaviors are guaranteed by constraints imposed by a deterministic finite-state automaton. The obtained results revealed that 99% of behaviors are well classified. Standing, lying on each side and changing between these positions are classified with 100% accuracy, followed by stationary with 99% sensitivity, 96% specificity, 99% precision, and 99% accuracy. The walking behavior is classified with 96% sensitivity, 99% specificity, 91% precision, and 98% accuracy.

GPS data is also an important source of information for behavior classification. In the work of Williams et al. (2019), they presented variable segmentation applied to GPS data gathered from 30 dairy cows at pasture. Using these segments, the performance of 13 ML algorithms (base learners) implemented in *Waikato Environment for Knowledge Analysis (WEKA)* were compared using default parameters in classifying grazing, resting and walking. Two stacking ensembles were then derived using the WEKA implementations. The first ensemble contained the best performing base learners. The second ensemble was an optimized version derived using a manual ensemble selection method. Both versions of the ensemble were evaluated on an independent test set derived from 10 cows. The ensembles performed well using base learners based on boosting algorithms: *Simple logistic (SL)*; *Logistic model trees (LMT)*, MLP, *Naive Bayes(NB)*, DT, SVM, *Naive Bayes tree (NBTree)*, *Logistic model trees (LMT)* and *Sequential minimal optimization (SMO)*.

Grass intake and rumination. Some researches have focused –specifically– on the classification of behavior related to grass intake and rumination. Andriamandroso et al. (2017) proposed an open algorithm, named *Statistical Model (SM)*, for the detection of cattle's grass intake and rumination activities. They mounted a smartphone on 19 grazing cows of different breeds, and recorded daily video sequences on the pasture of different forage. The final algorithm uses the average value and the standard deviation of two signals in a two-step discrimination tree: The gravitational acceleration on the x-axis (Gx) expressing the cows head movements, and the rotation rate on the same x-axis (Rx) expressing jaw movements. Validation on an independent database resulted in an average detection accuracy of 92% with a better detection for rumination (95%) than for grass intake (91%).

Another approach to the classification of actions, related to grass intake and rumination, is presented as follows. Chelotti et al. (2018) presented a new algorithm, called *Chew-Bite Intelligent Algorithm (CBIA)*. CBIA uses concepts and tools derived from pattern recognition and ML. CBIA includes: (i) a signal conditioning stage to attenuate the effects of noises and trends, (ii) a pre-processing stage to reduce the overall computational cost, (iii) an improved set of features to characterize jaw-movements, and (iv) a ML model to improve the discrimination capabilities of the algorithm. Three signal conditioning techniques and six ML models are evaluated. The overall performance is assessed on two independent data sets, using metrics like recognition rate, recall, precision and computational cost. The results demonstrated that CBIA achieves a 90% recognition rate, with a marginal increment of computational cost. Compared to the algorithms *Least Mean Squares (LMS)*, *Empirical Mode Decomposition (EMD)*, *Multilayer Perceptron (MLP)*, *Decision Trees (DT)* and SVM, CBIA improves the recognition rate by 10%, even in difficult scenarios.

The following work presents a different kind of study, in which they evaluate the effect of rumination and grass intake of different pastures. Alvarenga et al. (2020) developed a classifier of biting and chewing activities of sheep during grazing. Two studies were conducted. The first study evaluated the effect of two diverse pasture species on feeding behavior using micro-sward boxes: forage oats and perennial ryegrass. Two 4-year-old Merino ewes grazed each species, for approximately four, two-minute sessions, over two separate days, one week apart. In the second study, the effect of sward height was investigated using nine plots of ryegrass with three different sward heights, grazed by three 3-year old Merino ewes, for 10 min each. Forty-four features were calculated, from the acceleration signals, and used to classify behaviors using a DT algorithm to determine model accuracy, sensitivity, specificity, and precision. For the micro-sward study, bite activity was classified with a precision of 90.5% for the evaluation data set; whereas, for the validation data set, it was classified with a precision of 98.1% for the 5 s time interval. Accuracy of the DT model increased as time interval increased for both data sets.

There is research focusing –specifically – on the ingesting behavior for different grasses. Campos et al. (2018) presented a method to classify the ingesting behavior of ruminants by means of *Surface Electromyography (sEMG)* signal of the masseter muscle. Despite the similar properties of the grasses, the food-recognition results were reasonable. Feeding and rumination were discriminated with relatively high accuracy. The whole data acquisition, instrumentation and pattern recognition system follow the sequence outlined in Fig. 4. The evaluated classifiers methods applied to EMG pattern recognition are: LDA, QDA, SVM, *Multilayer-Perceptron Neural Network (MLP-NN)*, *Radial-Basis-Function Neural Network (RBF-NN)* and KNN. The MLP-NN was the classifier with the highest mean accuracy for almost every scenario and feature set.

In Fig. 4, the cattle are in rumination, and through the sEMG sensor, data is acquired. After, through signal processing, data is segmented. Such characteristics are created and a classification algorithm is used to be able to predict the animal's intake.

3.1.2. Animal monitoring

Documents focused on animal monitoring are divided into and general monitoring, monitoring calving and estrus, monitoring ingesting behavior, and muzzle-point recognition.

General monitoring. The following work presents a platform of general-purpose for animal monitoring. Debauche et al. (2019) proposed a lambda-cloud architecture coupled to a scientific sharing platform used to archive, and a process of high-frequency data processing is proposed to integrate future developments of the *Internet of Things (IoT)*, applied to the monitoring of domestic animals. An application to the study of cattle behavior on pasture-based on data recorded with the IMU of iPhone 4s is exemplified. The package comes with a web interface to encode the actual behavior observed on videos and to synchronize observations with the sensor signals.

There is general-purpose monitoring application based on computer vision and drones, proposed by Andrew et al. (2017). They explained that computer vision can use deep neural architectures, which are well-suited to perform automated Holstein Friesian cattle detection. They introduced a video processing pipeline composed of standard components, to efficiently process dynamic herd footage filmed by *Unmanned Aerial Vehicles (UAVs)*. They showed that Friesian cattle detection and localization can be performed robustly with an accuracy of 99.3% with this data. They evaluated the individual identification, exploiting coat uniqueness on 940 RGB stills taken after milking in-barn (89 individuals, accuracy = 86.1%). They also evaluated identification via a video processing pipeline on 46,430 frames (approx. 20 s length each) of UAV footage taken during grazing (23 individuals, accuracy = 98.1%).

Calving and estrus. For better management of animal health, it is required to know the exact time when calving and estrus occur. Benaissa et al. (2020) used *Logistic regression (LR)* models and data from calving and estrus sensors. The detection performance within different time intervals (24 h, 12 h, 8 h, 4 h, and 2 h) before calving was investigated. In general, for both calving and estrus, the performance of the detection, within 24 h, was lower than for 8 h-24 h. However, the use of a combination of sensors increased the performance for all investigated detection time intervals. For calving, similar results were obtained for

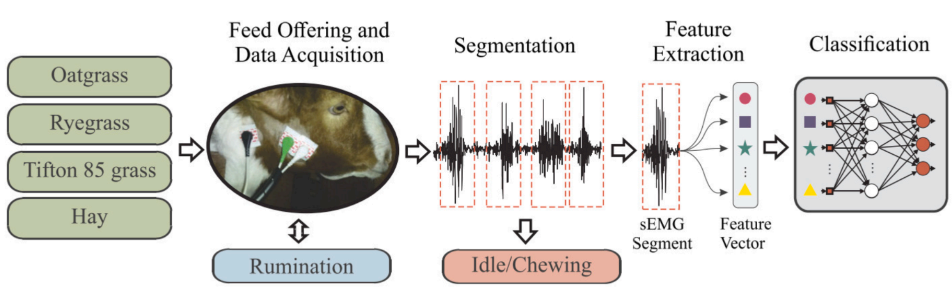


Fig. 4. Complete animal sEMG processing: data acquisition, signal segmentation, feature extraction and classification (Campos et al., 2018).

the detection within 24 h, 12 h, and 8 h. When one sensor was used for calving detection, within 24–8 h, the localization sensor performed best (Precision 73–77%, Sensitivity 57–58%, *Area under the receiver operating characteristic curve [AUC]* 90–91%), followed by the leg-mounted accelerometer (Precision 67–77%, Sensitivity 54–55%, AUC = 88–90%) and the neck-mounted accelerometer (Precision 50–53%, Sensitivity 47–48%, AUC = 86–88%). As for calving, the results of estrus were similar for the time intervals 24 h-8 h. In this case, similar results were obtained when using any of the three sensors, separately, as when combining neck and leg-mounted accelerometers (Precision 86–89%, Sensitivity 73–77%). For both calving and estrus, the performance improved when localization was combined with either, the neck or leg mounted accelerometers, especially, for sensitivity (73–91%).

Ingesting behavior. In monitoring, the goal is not to classify the action during grass intake or rumination, but also to estimate the amount of fiber intake. Campos et al. (2019) proposed a non-invasive method for fiber intake estimation on ruminants. Campos et al.'s method used a sEMG-based sensor system. In order to acquire bite/chewing sEMG signals, superficial disposable electrodes were placed on three uncastrated male goats' masseter muscle, housed in individual pens, and data was sampled during eating using an analog-to-digital converter. Feed samples and-left-overs were weighed before and after the experiment, respectively, to estimate the total intake. Electromyographic preprocessed data was sent to a computer, where seven signal features were extracted. Feed intake was modeled by fitting sEMG features as a predictor by means of a linear model. The results indicated that fiber intake could be successfully predicted in goats eating several forages (Tifton 68 and Tifton 85 grass hay, bad quality Tifton hay, and forage oat hay) with *a coefficient of determination (R2)* higher than 0.867, using a signal feature called *Slope Sign Change (SSC)* as a predictor. SSC expresses signal frequency characteristics and indicates the physiological aspects of bite/chewing.

Muzzle-point recognition. Kumar et al. (2017) have focused on a very particular topic of great interest in cattle. They proposed a muzzle-point recognition based on *Fisher locality*, using a projection algorithm for the recognition of cattle in real-time. They have captured images of animals, using a surveillance camera, and transferred them to the server by wireless network technology. The efficacy of the proposed muzzle-point recognition approach for cattle yields 96.87% recognition accuracy for identifying individual cattle. The proposed approach has a 10.25 s recognition time for enrollment and identification of individual cattle on different sizes of muzzle-point images.

3.2. *Machine learning for animal health in PLF*

In this section, the analysis of the documents is also divided into the classification of animal behavior and animal monitoring.

3.2.1. *Animal behavior*

The following documents focus on the classification of animal behavior. This section is divided into two parts. First, general works of animal-behavior classification for livestock and, after, some specific to sow's behavior are presented.

General-livestock behavior. Some works are focused on behavior classification for livestock in general. As an example, Dominiak and Kristensen (2017) proposed three methods to classify or prioritize alerts: *Fuzzy Logic (FL)*, *Naive Bayesian Network (NBN)* and *Markov Hidden Phase model.* NBN shows the greatest potential for future reduction of alerts from sensor-based detection models in livestock production.

Bishop et al. (2019) described another approach to classify behavior based on livestock vocalization. They proposed a multi-purpose livestock vocalization classification algorithm using audio-specific feature extraction techniques and ML models. A comparison of *Mel-Frequency Cepstral Coefficients (MFCCs)* and *Discrete Wavelet Transform (DWT)*-based features was conducted. Classification was determined using a SVM model. High accuracy was achieved for all data sets (sheeps: 99.29%, cattles: 95.78%, dogs: 99.67%). Computational timing reveals that the DWT-based features are faster; such features decrease by 14.81 – 15.38% execution time.

Sow's behavior. In order to control sow's health, an approach is to detect and classify sow's actions using images. Lao et al. (2016) described a computational algorithm for the analysis of depth images, and presented its performance in recognizing the sow's behaviors as compared to manual recognition. The images were acquired at 6 s intervals on three days of a 21-day lactation period. Based on the analysis of the 6 s interval images, the algorithm had the following accuracy in the classification: 99.9% in lying, 96.4% in sitting, 99.2% in standing, 78.1% in kneeling, 97.4% in feeding, 92.7% in drinking, and 63.9% in transitioning between behaviors. According to Lao et al., the lower classification accuracy, for the transitioning category, presumably, stemmed from insufficient frequency of the image acquisition.

Another approach to classify sow's behavior is using a Kinect sensor to identify sow's postures. Zheng et al. (2018) developed a detection system that consists of a Kinect v2 sensor that acquires depth images and a program that identifies sow postures. The depth images of the testing data set of a sow were acquired at 5 frames per second in 24hours on the 15th day of postpartum, and the training data set was collected by some different sows. Since the identification performance from RGB images are impacted by the color and illumination variations caused by an in situ heat lamp and day-night cycle, Zheng et al. showed that the automatic detection from depth images could avoid disturbances of the light. Zheng et al. (2018) found that the sow spent a greater amount of time in recumbency (92.9% at night and 84.1% during the daytime), as compared with standing (0.4% at night and 10.5% during the daytime) and sitting (0.55% at night and 3.4% during the daytime). According to Zheng et al., statistically, the sow's activity level is non-uniform in 24-h of a day, and its preferred lying positions are accordant with the pen's floor design.

3.2.2. *Animal monitoring*

For animal monitoring, the works are divided into three tasks. First, general works on animal monitoring are presented. After, works related to lameness monitoring are detailed. Finally, present works related to respiratory problems are introduced.

General monitoring. In the previous section, how vocalization can be used for behavior vocalization was presented. Vocalization can also be used for animal monitoring of thermal comfort. Du et al. (2020) developed a vocalization detection method, based on ML, to assess thermal-comfort condition. For extraction of the vocalizations, nine source-filters related temporal and spectral features were chosen, and a SVM-based classifier was developed. As a result, the classification performance of the SVM model was 95.1 ± 4.3% (the sensitivity parameter) and 97.6 ± 1.9% (the precision parameter). Based on the developed algorithm, Du et al. illustrated that a significant correlation exists between specific vocalizations (alarm and squawk call) and thermal comfort indices (temperature-humidity index, THI).

As another alternative to vocalization is to monitor movement and speed because such variables can explain an animal's well-being. Doulgerakis et al. (2019) presented an automated system with a single type of wireless sensor to record indicators of an animal's well-being: movement, speed and geolocation of the animal. Their system was made with a low implementation cost, based on *Deep Neural Networks (DNN)*-pattern-recognition algorithms. According to Doulgerakis et al., the solution provides end-users (farmers) with usable and effective information visualization, so that they take proper actions.

Deep learning has shown to be a technique very useful in animal monitoring. On the one hand, Fonseca et al. (2019) developed a model to predict stress in piglets, monitoring skin temperature using an infrared camera. A total of 40 piglets (20 males and 20 females), from 1 to 22 weeks, under different stress conditions, had the skin temperature recorded during the farrowing and nursery phases. The stresses studied were hunger, pain, thirst, heat/cold, and the normal. The attributes considered, in the analysis were classified using data mining, with stress condition as the target. They used a DT classifier to predict the stress condition, and found that the pain and thirst attributes had better precision, in the model, with the maximum-surface temperature and the sex of the pig.

Another work using deep learning is that of Cowton et al. (2019). They combined a deep-CNN object-localization method, named R-CNN, with two potential real-time multi-object tracking methods, to create a system that can –autonomously– localize and track individual pigs. Cowtom's et al. captured data using RGB cameras. Their system is able to: (i) localize pigs in individual frames, with 0.901 mean average precision, (ii) track individual pigs across video footage, with 92% multi-object tracking accuracy, and (iii) re-identify them, after occlusions and dropped frames, with 0.862 mean average precision.

Deep learning has shown a particular success when using 3D scanning. Sousa et al. (2018) built a sensor array consisting of a two-dimensional (2D) laser scanner and an encoder for the three-dimensional (3D) scanning of the back area of 107 Nellore cattle. The scanning data were taken in four periods, every 28 days, which allowed generating 304 clouds of points. They built an algorithm using delaunay and convex hull methods to obtain the height of the cattle's rump and the area of the rear view. They also built a neural model, with a multi-layered architecture. The input of the neural model was the rump height and rear view, and the output was an estimation of the live weight. Model performance was evaluated by comparing the estimated and measured weight of the cattle using linear-regression parameters. The coefficient of determination was 0.85 and the mean absolute percentage error was 4.57%, in previously unseen data.

In addition to machine learning, there is also research ondata visualization of the animal's well-being. Van Van Hertem et al. (2017) presented the development of a web-based tool for data visualization. Data was collected from five broiler farms and ten pig farms across Europe. At the farms, a number of variables were automatically measured including climate data, production data, environmental data, and data about animal behavior coming from cameras and microphones. Simultaneously, the health of the animals was assessed by trained assessors using a standardized health quality protocol. All data was gathered, stored and processed on a daily basis. End-users of the tool were trained on how to interpret the available information on the visualization tool.

Monitoring has also been applied to poultry farming. Stefanova (2017), Stefanova (2019) described the realization of a precision-livestock management-software that delivers monitoring and collaborative capabilities to improve laying hens health at industrial poultry farms. The online platform as a zoo-technical diary connects egg and breeding farms through cloud technologies to provide continuous data-recording, automatic comparisons between actual and expected production indicators, and integrated data-flow between the parties. According to Stefanova et al., breeding farms benefit from enhanced competitiveness, improved supplier-client relationships, while egg farms enjoy management precision, timely feedback on animals' health and economic benefits.

In the last work on general monitoring, Makinde et al. (2019) discussed the opportunities to apply research and methodologies emerging from *animal-computer interaction (ACI)* to help improve the usability and overall utility of PLF technologies for both, human and animal users.

Heat Monitoring. Heat stress affects animal health and productivity. In one work, Liu et al. (2018) developed a method for turkey-sound analysis and determine whether heat stress can induce specific turkey vocalizations. Forty-eight turkeys were bred in eight isolated rooms, under laboratory conditions, for eight weeks, and –randomly– allotted to heat-stress rooms and control rooms. SVM was used as the classifier for sound-signal recognition. The accuracy of classification was 88.75%. They demonstrated the possibility of using turkey vocal sound monitoring and analysis as an early warning tool for heat-stress detection. In another work, Milan et al. (2019) determined optimum supplemental heat requirements (supplied by heating lamps), for piglets, based on energy balance as a function of air temperature and animal body weight. They also defined the zone of least thermo-regulation of piglets, for a given weight, when supplemental heat is not provided. Energy balance was calculated using an ensemble of mechanistic models of bio-heat transfer that predicts hair-coat temperature, skin temperature, and skin-heat flux. Input temperatures were predicted from measured air temperature in the pen and supplemental heat using machine learning techniques. Predicted optimum supplemental heat showed an exponential-decay trend with increasing air temperature and/or animal weight.

Insemination Monitoring. Labrecque and Rivest (2018) proposed a novel approach to determine an optimal timing for insemination in sows, based on behavior analysis in real-time. Real-time analysis allows finding the statistically-optimal timing for insemination based on behavior-pattern recognition. The system was used for 21 months, in seven commercial farms, for 20,485 weaned sows. When used as a complement to a daily heat detection, the system allowed to easily manage 99.78% of the sows.

Lameness monitoring. Lameness indicates that something is wrong with the cattle. An approach to detect lameness is to use a sensor to track leg movement. Zhao et al. (2018) analyzed leg swing using computer-vision techniques, and developed an automatic and continuous system for scoring the locomotion of cows to detect and predict lameness with high accuracy. The focus was the quantification of the movement pattern of cows and the demonstration of the possibility of classifying lameness, using the features extracted from movement analysis. Side-view videos were recorded after the cows were milked. Cows were scored by an expert on a scale from 1 (healthy) to 3 (severely lame). The data set included 621 videos from 98 cows. The motion curve was plotted by extracting the position of the moving leg by image processing, and the motion curve was analyzed to generate six features referring to the gait asymmetry, speed, tracking up, stance time, stride length, and tenderness. The DT classifier was applied to the data set, and 2-, 3-, and 10-fold cross-validation was used to verify the performance of the algorithm. The accuracy of the classification was 90.18%, and the average sensitivity and specificity were 90.25% and 94.74%, respectively.

Another approach to detect lameness is to use *quantile regression forests (QRT)* instead of DT. Diez-Olivan et al. (2019) created a decision-support system based on environmental indicators, and on the weights, leg problems and mortality rates. The data-driven modeling process is performed by the QRT approach that allows estimating growth, health and mortality parameters on the basis of environmental deviations from optimal farm conditions.

Respiratory problems. The following works study respiratory diseases for growing pigs. Cowton et al. (2018) designed and evaluated a deep-learning-based methodology for animal-health monitoring. The methodology is, specifically, for the early detection of respiratory disease, in growing pigs, based on environmental sensor data. Two *recurrent neural networks (RNNs)* were used to create an *autoencoder (GRU-AE)*. The models received environmental data, collected from a variety of sensors, to detect anomalies. Cowton et al. used *Particle Swarm Optimization (PSO)* to raise alerts. The results showed that a change in the environment can result in pigs showing symptoms of respiratory disease within 1–7 days, meaning that there is a period of time during which their keepers can act to mitigate the negative effect of respiratory diseases, such as *porcine reproductive and respiratory syndrome (PRRS)*, a common and destructive disease endemic in pigs.

Health status. *FAffa MAlan CHArt (FAMACHA)* is a scoring system for detecting levels of anemia caused by parasite infection in small ruminants. FAMACHA is an effective way to identify individuals needing treatment. Montout et al. (2020) used training/test data sets, using FAMACHA scores, to construct a SVM classifier, to predict individuals that needed treatment.

4. Discussion

In this section, a summary of the revisions, the limitations of the reviewed works, and challenges for future works are presented.

4.1. Review summary

In this review, we have found that grazing, using ML techniques, in recent years, has focused –mainly– on sheep (40%) and cattle (60%). The most commonly used ML techniques for grazing are classification techniques. Most works on grazing have focused on obtaining, analyzing and understanding animal variables. With respect to animal health, a recent work has focused on how to improve animal production by controlling their health. Research has shown that estimating food consumption and monitoring feeding behavior are key activities to assess the health of animals. Also, most research on animal health has focused on understanding animal behavior. Nonetheless, both animal monitoring and animal behavior are complementary and needed to measure animal health and have not been considered together.

Fig. 5 summarizes all 35 studies, examining the problem they addressed, the proposed solution, and the tools and techniques used with the best performance.

Fig. 6 shows that the most commonly used ML techniques are SVM and DT. There are some possible explanations. For example, SVM has had great acceptance because it has the ability to discriminate data of different kinds by the means of kernels: The theoretical basis of kernels used in SVM makes it an exceptional tool for generalization in complex problems (Lemaire et al., 2018). On the other hand, DT predicts the value of a target variable by learning simple decision rules inferred from the characteristics of the data: This makes it easy to understand and interpret, requires little data preparation, and it is a white box model.

Fig. 7 shows the different sensors used in PLF. This figure shows the interest of the scientific community: In the first place, accelerometers and, in the second place, collar sensors. The third most used sensor is the video capture.

Accelerometers have been used in cattle to evaluate the activity patterns of dairy cattle (de Passillé et al., 2010), the resting behavior of calves (Bonk et al., 2013), and the analysis of the grazing vs. non-grazing in dairy cows (Nielsen, 2013). Accelerometers can be worn on a collar located on specific parts of the animal's body. A collar is a very important instrument because it can have all kinds of sensors, which are in charge of collecting the data and sending data, such as the GPS sensor. Video capture is the easiest way to monitor animals. By installing the cameras, it is possible to obtain data about what is happening to the animal and its environment in a non-invasive way.

4.2. Limitations of reviewed works

In this section, the limitations of ML for grazing and animal health in PLF are presented.

Machine Learning for Grazing in PLF. Management of pastures allows maintaining production levels in terms of forage produced per hectare, and milk and meat production per hectare. Therefore, analyzing and preparing the soil where the pasture will be established will give a good pasture management. In addition, a good pasture allows extensive livestock to obtain their food from grazing. In exchange, livestock fertilizes the land and facilitates soil tillage, demonstrating the complementary relationship between livestock and pasture agriculture (Lemaire et al., 2018).

In spite of the well-known importance of management of pastures, ML for grazing has focused, mostly, on classical classification techniques to understand the behavior of the animal. The use of these variables is centered on the animal; however, variables such as forage digestibility and animal production are not being used. Even worse, previous works have not taken into account soil variables, which, together with the animal activity, could lead to the creation of robust models to improve grazing methods, and, as a consequence, to a higher economical production. Finally, there are no works using unsupervised learning, or new ML techniques, such as semi-supervised learning (Cerrada et al., 2019), inherited learning (Salakhutdinov et al., 2012; Puerto and Aguilar, 2016) and meta-learning (Finn et al., 2017).

Machine Learning for Animal health in PLF. On ML for animal

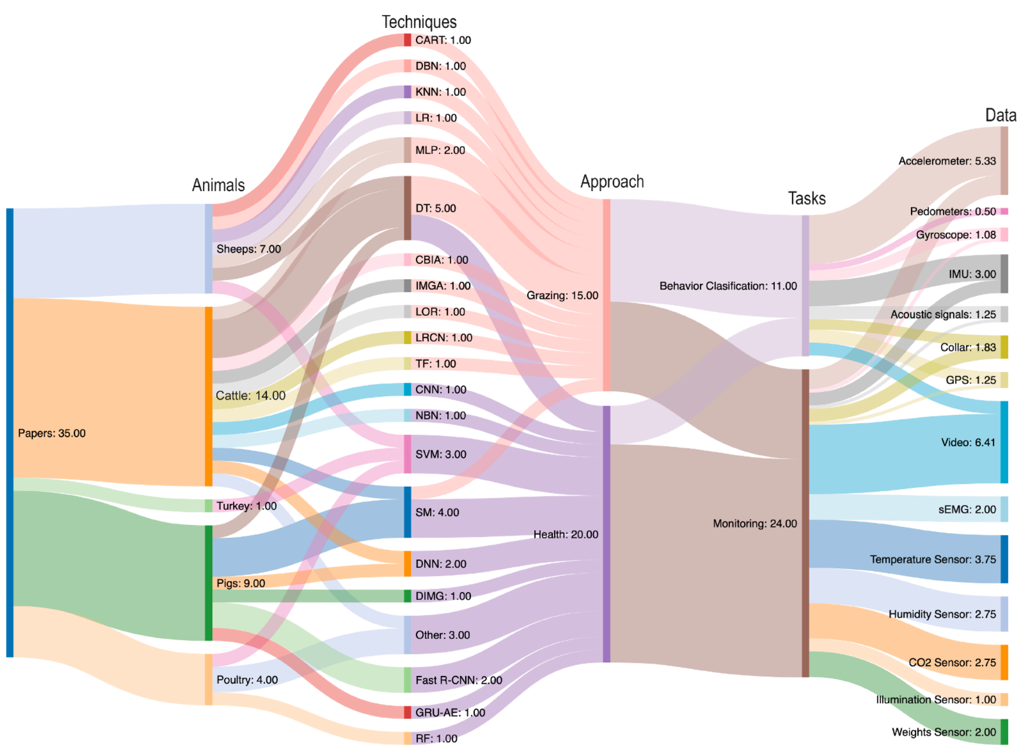


Fig. 5. Trends of reviewed documents.

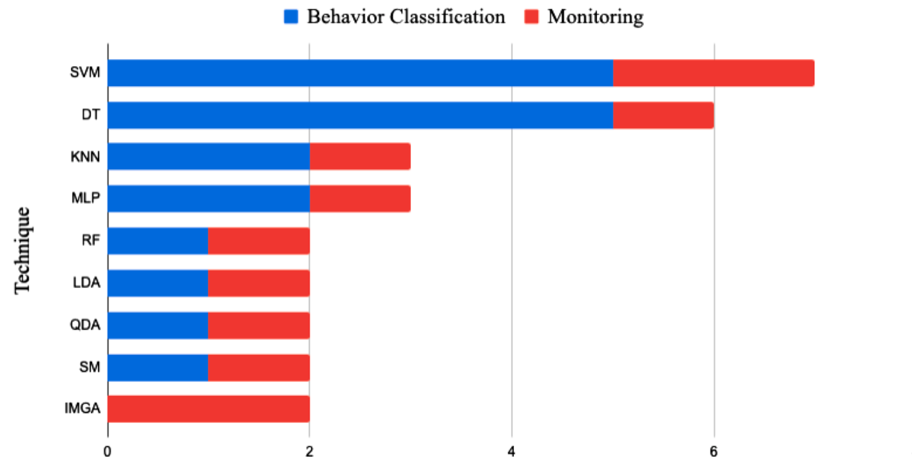


Fig. 6. Most used ML techniques in PLF.

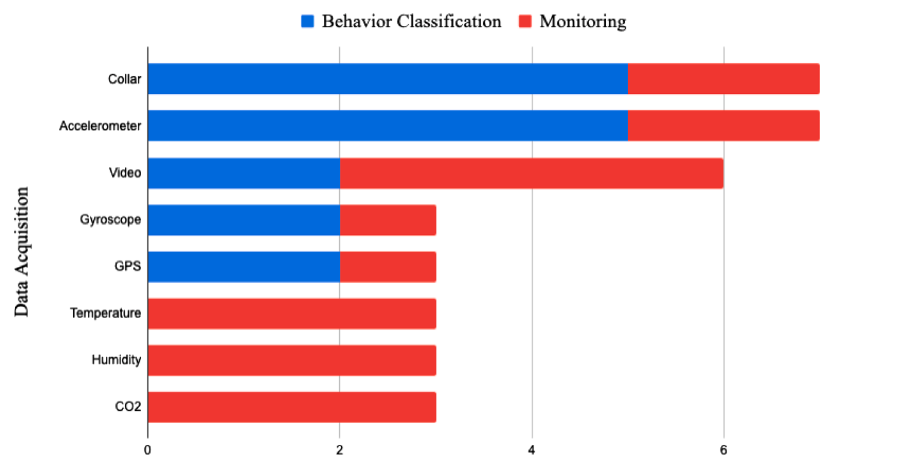


Fig. 7. Most used data-acquisition technologies in PLF.

health, the works have been centered on monitoring the animal and, also, on classic ML techniques for classification. There are no works where the degree of animal comfort under its environment is analyzed, except two on thermal comfort. There are no works of ML for diagnosis of diseases, prediction or prescription of treatments, except for two works on the diagnosis of lameness. Neither there are works in which one brings together data of animal health with the state of water drinkers. Dirty water can contain unhealthy bacteria for animals, and, eventually, dirty water could contribute to the reduction of production.

4.3. Challenges

Challenges for future works can be, broadly, classified into three categories: Machine learning, sensors and data sources, and data management. In what follows, each category of challenges is presented.

Machine learning. There is no evidence that grazing and animal health have been brought together in a ML model; therefore, trying to combine these aspects results in an interesting challenge. Bringing together grazing and animal health will also allow developing models of diagnosis and prescription as tools for the prevention and control of animal diseases. Such tools will entail considerable benefits to livestock production, food safety, public health, animal health and access to international markets.

To bring together grazing and animal health issues, a recommendation is the implementation of federated learning (Yang et al., 2019; Li et al., 2020). A federated ML model, for grazing and animal health, will allow data sets from different farms to be interconnected so that ML models work together, and they can be adapted to changing environments and event trends. Finally, to give autonomy to PLF, it is necessary to explore concepts such as *autonomous cycles of data analysis tasks* that consider all ML tasks working together (Aguilar et al., 2018b; Aguilar et al., 2018a), and often include meta-learning (Finn et al., 2017). This last challenge can be extended with concepts from the area of multi-agent systems (Aguilar et al., 2007), which allow modeling the entire PLF context in a natural way.

Another challenge in ML is about how to improve the classification performance of the algorithms used in PLF. Williams et al. (2019) determined that ensemble algorithms are very promising in PLF, but more work is needed to explore different strategies available, in the literature, for PLF tasks. Such PLF tasks can benefit from the added power of ensemble learning, which is already successful in other areas. As a final recommendation, many tasks can be treated not only as a classification problem, but also as a semi-supervised or a meta-learning problem.

Sensors and data sources. Systems based on sensor technologies are called upon to provide decisive information on conditions influencing livestock; for instance, to measure humidity, temperature, wind, rain, soil quality and animal health status. The use of sensors and data analysis focusing on real farm problems will ensure that farms are environmentally sustainable, and avoid wastes that have a major impact on soil contamination. The increasing availability of large data sources and data sets will encourage more initiatives, projects and new ventures in the livestock sector.

To improve ML for grazing, there are three key factors for pasture management: the frequency, intensity and duration of grazing. Nonetheless, these variables are not being taken into account. According to Arcos et al. (2019), considering such variables would help to correctly define the condition of the pasture, before and after grazing, so that grazing takes place at the right time and with the right intensity (Bell et al., 2014).

In ML for animal health, previous works have focused on pigs (31%), sheep (8%), poultry (38%) and beef (23%), and have a greater focus on animal monitoring with 75%, and behavioral classification with 25%. On animal health and environmental measurement sensors, the most used sensors are temperature, humidity, and CO_2 sensors. Van Hertem et al. (2017) proposed new sensors for the detection of early-signs of deviations in animal health, with low financial and energy costs. A challenge, in addition to the development of sensors, is how to determine the relevant signals: This is a problem of feature engineering.

Data management. There are some challenges related to data management in PLF. Stefanova (2019) proposed that most farm data is found in notebooks or in zoo-technical information systems. A challenge is the automatic digitization of all the useful data of a farm for the context of PLF.

Once farmers achieve automatic digitization, there is a challenge on how to make the system more energy efficient. To improve energy-saving, farmers have to merge data from different sources, reduce dimensions, minimize energy consumption in on-farm equipment, and compress data. In fact, Debauche et al. (2019) concluded that the use of data-compression algorithms can optimize battery-power consumption in the context of PLF. Another issue to save energy in PLF is how to minimize the amount of data transmitted to the cloud, as proposed by Doulgerakis et al. (2019). Doulgerakis et al. also proposed to apply intelligent energy-saving algorithms to optimize battery life.

A final challenge in data management is the following: Van Hertem et al. (2017) determined that the main barrier among farmers living in rural areas is the appropriate data visualization. For this reason, Van Hermet et al. concluded that a user-friendly interface, for the correct presentation of data, is a major challenge.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Rodrigo García is supported by a PhD grant financed by Universidad EAFIT. All the authors would like to thank the “Vicerrectoría de Descubrimiento y Creación”, of Universidad EAFIT, for their support on this research.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.compag.2020.105826.

References

Achour, B., Belkadi, M., Aoudjit, R., Laghrouche, M., 2019. Unsupervised automated monitoring of dairy cows' behavior based on Inertial Measurement Unit attached to their back. Comput. Electron. Agric. 167.

Aguilar, J., Cerrada, M., Hidrobo, F., 2007. A methodology to specify multiagent systems. In: Nguyen, N., Grzech, A., Howlett, R., Jain, L. (Eds.), Agent and Multi-Agent Systems: Technologies and Applications. Springer, Berlin Heidelberg, pp. 92–101.

Aguilar, J., Cordero, J., Buendia, O., 2018a. Specification of the autonomic cycles of learning analytic tasks for a smart classroom. J. Educat. Comput. Res. 56, 866–891.

Aguilar, J., Sánchez, M., Cordero, J., Valdiviezo-Diaz, P., Rodrigo, L., Chamba-Eras, L., 2018b. Learning analytics tasks as services in smart classrooms. Univ. Access Inf. Soc. 17, 693–709.

Alvarenga, F.A.P., Borges, I., Oddy, V.H., Dobos, R.C., 2020. Discrimination of biting and chewing behaviour in sheep using a tri-axial accelerometer. Comput. Electron. Agric. 168.

Andrew, W., Greatwood, C., Burghardt, T., 2017. Visual localisation and individual identification of holstein friesian cattle via deep learning. In: 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), pp. 2850–2859.

Andriamandroso, A.L.H., Lebeau, F., Beckers, Y., Froidmont, E., Dufrasne, I., Heinesch, B., Dumortier, P., Blanchy, G., Blaise, Y., Bindelle, J., 2017. Development of an open-source algorithm based on inertial measurement units (IMU) of a smartphone to detect cattle grass intake and ruminating behaviors. Comput. Electron. Agric. 139, 126–137.

Arcos, C., Lascano, P., Guevara, R., 2019. Manejo de asociaciones gramíneas-leguminosas en pastoreo con rumiantes para mejorar su persistencia, la productividad animal y el impacto ambiental en los trópicos y regiones templadas. Revista Ecuatoriana de Ciencia Animal 2 (2), 1–31.

Astill, J., Dara, R.A., Fraser, E.D.G., Roberts, B., Sharif, S., 2020. Smart poultry management: Smart sensors, big data, and the internet of things. Comput. Electron. Agric. 170.

Bahlo, C., Dahlhaus, P., Thompson, H., Trotter, M., 2019. The role of interoperable data standards in precision livestock farming in extensive livestock systems: A review. Comput. Electron. Agric. 156, 459–466.

Banhazi, T.M., Black, J.L., 2009. Precision livestock farming: a suite of electronic systems to ensure the application of best practice management on livestock farms. Austral. J. Multi-Discip. Eng. 7 (1), 1–14.

Banhazi, T.M., Lehr, H., Black, J.L., Crabtree, H., Schofield, P., Tscharke, M., Berckmans, D., 2012. Precision Livestock Farming: An international review of scientific and commercial aspects. Int. J. Agric. Biol. Eng. 5 (3), 1–9.

Bell, L., Moore, A., Kirkegaard, J., 2014. Evolution in crop-livestock integration systems that improve farm productivity and environmental performance in Australia. Eur. J. Agron. 57, 10–20.

Benaissa, S., Tuyttens, F.A.M., Plets, D., Trogh, J., Martens, L., Vandaele, L., Joseph, W., Sonck, B., 2020. Calving and estrus detection in dairy cattle using a combination of indoor localization and accelerometer sensors. Comput. Electron. Agric. 168.

Benjamin, Madonna, Yik, Steven, 2019. Precision livestock farming in swine welfare: a review for swine practitioners. Animals 9 (4), 133.

Berckmans, D., 2014. Precision livestock farming technologies for welfare management in intensive livestock systems. Revue Sci. Tech. 33, 189–196.

Bishop, J.C., Falzon, G., Trotter, M., Kwan, P., Meek, P.D., 2019. Livestock vocalisation classification in farm soundscapes. Comput. Electron. Agric. 162, 531–542.

Bonk, S., Burfeind, O., Suthar, V.S., Heuwieser, W., 2013. Technical note: Evaluation of data loggers for measuring lying behavior in dairy calves. J. Dairy Sci. 96 (5), 3265–3271.

Caja, G., Castro-Costa, A., Knight, C., 2016. Engineering to support wellbeing of dairy animals. J. Dairy Res. 83 (2), 136–147.

Campos, D.P., Abatti, P.J., Bertotti, F.L., Hill, J.A.G., da Silveira, A.L.F., 2018. Surface electromyography segmentation and feature extraction for ingestive behavior recognition in ruminants. Comput. Electron. Agric. 153, 325–333.

Campos, D.P., Abatti, P.J., Bertotti, F.L., De Paula Vieira, A., Hill, J.A.G., da Silveira, A.L. F., 2019. Short-term fibre intake estimation in goats using surface electromyography of the masseter muscle. Biosyst. Eng. 183, 209–220.

Cerrada, M., Aguilar, J., Altamiranda, J., Sánchez, R., 2019. A hybrid heuristic algorithm for evolving models in simultaneous scenarios of classification and clustering. Knowl. Inf. Syst. 61, 755–798.

Chelotti, J.O., Vanrell, S.R., Galli, J.R., Giovanini, L.L., Rufiner, H.L., 2018. A pattern recognition approach for detecting and classifying jaw movements in grazing cattle. Comput. Electron. Agric. 145, 83–91.

Cowton, J., Kyriazakis, I., Plötz, T., Bacardit, J., 2018. A combined deep learning GRU-autoencoder for the early detection of respiratory disease in pigs using multiple environmental sensors. Sensors 18 (8).

Cowton, J., Kyriazakis, I., Bacardit, J., 2019. Automated individual pig localisation, tracking and behaviour metric extraction using deep learning. IEEE Access 7, 108049–108060.

Debauche, O., Mahmoudi, S., Andriamandroso, A.L.H., Manneback, P., Bindelle, J., Lebeau, F., 2019. Cloud services integration for farm animals' behavior studies based on smartphones as activity sensors. J. Ambient Intell. Humanized Comput. 10 (12), 4651–4662.

de Passillé, A.M., Jensen, M.B., Chapinal, N., Rushen, J., 2010. Technical note: Use of accelerometers to describe gait patterns in dairy calves. J. Dairy Sci. 93 (7), 3287–3293.

Diez-Olivan, A., Averós, X., Sanz, R., Sierra, B., Estevez, I., 2019. Quantile regression forests-based modeling and environmental indicators for decision support in broiler farming. Comput. Electron. Agric. 161, 141–150.

Dominiak, K.N., Kristensen, A.R., 2017. Prioritizing alarms from sensor-based detection models in livestock production - A review on model performance and alarm reducing methods. Comput. Electron. Agric. 133, 46–67.

Doulgerakis, V., Kalyvas, D., Bocaj, E., Giannousis, C., Feidakis, M., Laliotis, G.P., Patrikakis, C., Bizelis, I., 2019. An animal welfare platform for extensive livestock production systems. In: CEUR Workshop Proceedings, vol. 2492.

Du, X., Carpentier, L., Teng, G., Liu, M., Wang, C., Norton, T., 2020. Assessment of laying hens' thermal comfort using sound technology. Sensors 20 (2).

Espinosa, H.R., Duque, A.B., Arroyave, J.E.N., 2016. Application of geographic information systems (gis) for the implementation of precision farming. Livestock Res. Rural Develop. 28 (8).

Finn, C., Abbeel, P., Levine, S., 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of the 34th International Conference on Machine Learning - Volume 70, ICML'17, pp. 1126–1135.

Fogarty, E.S., Swain, D.L., Cronin, G.M., Moraes, L.E., Trotter, M., 2020. Behaviour classification of extensively grazed sheep using machine learning. Comput. Electron. Agric. 169.

Fonseca, Felipe N., Abe, Jair M., Nääs, Irenilza A., Cordeiro, Alexandra F.S., Amaral, Fabio V., 2019. Automatic prediction of stress in piglets using skin temperature. In: 2019 ASABE Annual International Meeting. American Society of Agricultural and Biological Engineers, p. 1.

Kumar, S., Singh, S.K., Singh, R.S., Singh, A.K., Tiwari, S., 2017. Real-time recognition of cattle using animal biometrics. J. Real-Time Image Proc. 13 (3), 505–526.

Labrecque, Jacquelin, Rivest, Joël, 2018. A real-time sow behavior analysis system to predict an optimal timing for insemination. In: 10th International Livestock Environment Symposium. American Society of Agricultural and Biological Engineers, p. 1.

Lao, F., Brown-Brandl, T., Stinn, J.P., Liu, K., Teng, G., Xin, H., 2016. Automatic recognition of lactating sow behaviors through depth image processing. Comput. Electron. Agric. 125, 56–62.

Lemaire, G., Giroud, B., Bathily, B., Lecomte, P., Corniaux, C., 2018. Toward integrated crop-livestock systems in west africa: A project for dairy production along senegal river. Agroecosyst. Diversity: Reconcil. Contemp. Agric. Environ. Qual. 275–285.

Li, T., Sahu, A.K., Talwalkar, A., Smith, V., 2020. Federated learning: Challenges, methods, and future directions. IEEE Signal Process. Mag. 37 (3), 50–60.

Liu Longshen, Ni Ji-Qin, Li Yansen, Erasmus Marisa, Stevenson Rachel, Shen Mingxia, 2018. Assessment of heat stress in turkeys using animal vocalization analysis. In: 2018 ASABE Annual International Meeting. American Society of Agricultural and Biological Engineers, p. 1.

Machado, Miguel, Silva, Iran, 2019. Does farm animals experience emotions and feelings? J. Animal Behav. Biometeorol. 7, 170–175.

Makinde, A., Islam, M.M., Scott, S.D., 2019. Opportunities for ACI in PLF: Applying animal- And user-centred design to precision livestock farming. In: ACM International Conference Proceeding Series. Association for Computing Machinery.

Mcloughlin, M.P., Stewart, R., McElligott, A.G., 2019. Automated bioacoustics: Methods in ecology and conservation and their potential for animal welfare monitoring. J. Roy. Soc. Interf. 16 (155).

Milan, Hugo F.M., Perano, Kristen M., Gebremedhin, Kifle G., 2018. Survey and future prospects in precision dairy farming. In: 10th International Livestock Environment Symposium. American Society of Agricultural and Biological Engineers, p. 1.

Milan, Hugo F.M., Maia, Alex Sandro Campos, Gebremedhin, Kifle G., 2019. Prediction of optimum supplemental heat for piglets. Trans. ASABE 62 (2), 321–342.

Montout Axel X., Bhamber Ranjeet S., Lange Debbie S., Ndlovu Doreen Z., Morgan Eric R., Ioannou Christos C., Terrill Thomas H., Van Wyk Jan A., Burghardt Tilo, Dowsey, Andrew W., 2020. Accurate and interpretable prediction of poor health in small ruminants with accelerometers and machine learning. bioRxiv.

Nielsen, P., 2013. Automatic registration of grazing behaviour in dairy cows using 3d activity loggers. Appl. Animal Behav. Sci. 148 (3), 179–184.

Norton, T., Chen, C., Larsen, M.L.V., Berckmans, D., 2019. Review: Precision livestock farming: building 'digital representations' to bring the animals closer to the farmer. Animal 13 (12), 3009–3017.

O'Grady, M., O'Hare, G., 2017. Modelling the smart farm. Informat. Process. Agric. 4 (3), 179–187.

Overgaauw, P., Vinke, C., Hagen, M.n., Lipman, L., 2020. A One Health Perspective on the Human-Companion Animal Relationship with Emphasis on Zoonotic Aspects. Int. J. Environ. Res. Public Health 17 (11), 1–29.

Perakis, K., Lampathaki, F., Nikas, K., Georgiou, Y., Marko, O., Maselyne, J., 2020. CYBELE – Fostering precision agriculture & livestock farming through secure access to large-scale HPC enabled virtual industrial experimentation environments fostering scalable big data analytics. Comput. Netw. 168.

Pomar, J., López, V., Pomar, C., 2011. Agent-based simulation framework for virtual prototyping of advanced livestock precision feeding systems. Comput. Electron. Agric. 78 (1), 88–97.

Porter, S., 2019. Production-phase greenhouse gases embedded within food loss and waste: magnitude, drivers, and mitigation potential. PhD thesis, University of Edinburgh.

Puerto, E., Aguilar, J., 2016. Learning algorithm for the recursive pattern recognition model. Appl. Artif. Intell. 30 (7), 662–678.

Salakhutdinov, R., Tenenbaum, J., Torralba, A., 2012. One-shot learning with a hierarchical nonparametric bayesian model. In: ICML Unsupervised and Transfer Learning.

Shine, P., Upton, J., Sefeedpari, P., Murphy, M.D., 2020. Energy consumption on dairy farms: A review of monitoring, prediction modelling, and analyses. Energies 13 (5).

Singer, R., Porter, L., Thomson, D., Gage, M., Beaudoin, A., Wishnie, J., 2019. Potential impacts on animal health and welfare of raising animals without antibiotics. bioRxiv.

Sousa, Rafael V., Tabile, Rubens A., Inamasu, Ricardo Y., Martello, Luciane S., 2018. Evaluating a lidar sensor and artificial neural network based-model to estimate cattle live weight. In: 10th International Livestock Environment Symposium. American Society of Agricultural and Biological Engineers, p. 1.

Stefanova, M., 2017. Precision poultry farming: Monitor and collaborate on health and welfare of laying hens through online zootechnical diary. In: Bournaris, T., Salampasis, M., Theodoridis, A. (Eds.), CEUR Workshop Proceedings, vol. 2030, CEUR-WS, pp. 39–45.

Stefanova, M., 2019. Precision poultry farming: software architecture framework and online zootechnical diary for monitoring and collaborating on hens' health. Commun. Comput. Informat. Sci. 953, 191–205.

Suparwito, H., Wong, K.W., Xie, H., Rai, S., Thomas, D., 2019. A Hierarchical Classification Method Used to Classify Livestock Behaviour from Sensor Data. Lecture Notes in Computer Science, 11909 LNAI, pp. 204–215.

Suryawanshi, K., Redpath, S., Bhatnagar, Y., Ramakrishnan, U., Chaturvedi, V., Smout, S., Mishra, C., 2017. Impact of wild prey availability on livestock predation by snow leopards. Roy. Soc. Open Sci. 4 (6).

Torres-Carrion, P., Gonzalez-Gonzalez Carina S., Aciar, S., Rodriguez-Morales, G., 2018. Methodology for systematic literature review applied to engineering and education. In: IEEE Global Engineering Education Conference, EDUCON, 2018-April, pp. 1364–1373.

Van der Burg, S., Bogaardt, M., Wolfert, S., 2019. Ethics of smart farming: Current questions and directions for responsible innovation towards the future. NJAS - Wageningen J. Life Sci. 90–91 (March 2020), 1–10.

Van Hertem, T., Rooijakkers, L., Berckmans, D., Peña Fernández, A., Norton, T., Berckmans, D., Vranken, E., 2017. Appropriate data visualisation is key to Precision Livestock Farming acceptance. Comput. Electron. Agric. 138, 1–10.

Vázquez D., J.A., Barker, Z.E., Hodges, H.R., Amory, J.R., Croft, D.P., Bell, N.J., Codling, E.A., 2015. Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system. Animal Biotelemet. 3 (1).

Vázquez-Diosdado, J.A., Paul, V., Ellis, K.A., Coates, D., Loomba, R., Kaler, J., 2019. A combined offline and online algorithm for real-time and long-term classification of sheep behaviour: Novel approach for precision livestock farming. Sensors 19 (14).

Vranken, E., Berckmans, D., 2017. Precision livestock farming for pigs. Animal Frontiers 7 (1), 32–37.

Williams, M.L., James, W.P., Rose, M.T., 2019. Variable segmentation and ensemble classifiers for predicting dairy cow behaviour. Biosyst. Eng. 178, 156–167.

Yang, Q., Liu, Y., Chen, T., Tong, Y., 2019. Federated machine learning: Concept and applications. ACM Trans. Intell. Syst. Technol. 10 (2).

Zhao, K., Bewley, J.M., He, D., Jin, X., 2018. Automatic lameness detection in dairy cattle based on leg swing analysis with an image processing technique. Comput. Electron. Agric. 148, 226–236.

Zheng, C., Zhu, X., Yang, X., Wang, L., Tu, S., Xue, Y., 2018. Automatic recognition of lactating sow postures from depth images by deep learning detector. Comput. Electron. Agric. 147, 51–63.

Appendix B

Autonomic computing in a beef-production process for precision livestock farming

Journal of Industrial Information Integration 31 (2023) 100425

Contents lists available at ScienceDirect

Journal of Industrial Information Integration

journal homepage: www.elsevier.com/locate/jii

Full length article

Autonomic computing in a beef-production process for Precision Livestock Farming

Rodrigo García [a,b], Jose Aguilar [a,d,f,h,*], Mauricio Toro [a], Nelson Pérez [e], Angel Pinto [c], Paul Rodríguez [c,g]

[a] GIDITIC, Universidad EAFIT, Medellín, Colombia
[b] Faculty of Engineering, Universidad del Córdoba, Montería, Colombia
[c] Faculty of Engineering, Universidad del Sinú, Montería, Colombia
[d] CEMISID, Universidad de Los Andes, Merida, Venezuela
[e] Grupo de Investigación de Telecomunicaciones, ULA, Mérida, Venezuela
[f] Universidad de Alcalá, Dpto. de Automática, Spain
[g] "El Rosario" Cattle Farm, Montería, Colombia
[h] IMDEA Networks Institute, Leganés, Madrid, Spain

ARTICLE INFO

Keywords:
Precision livestock farming
Autonomic computing
Industry 4.0
Artificial intelligence
Beef production
Data analytics

ABSTRACT

Precision livestock farming (PLF) offers farmers real-time monitoring and management system. PLF provides a real-time warning when something goes wrong so that the farmer can take immediate action to solve the problem. PLF introduces many new challenges and questions that must be resolved. Some of these challenges are related to the integration of grazing and animal health into the beef-production process. This article introduces an architecture for the self-managing of a beef-production farm. In particular, the architecture includes three autonomous cycles of data analysis tasks (ACODAT) that allow beef producers to have adequate coordination, optimization and planning of the productive process, which are: (i) circuit preparation, (ii) animal purchase, and (iii) animal fattening. This article also instantiates, in a farm, the autonomous animal-fattening cycle, as the first step towards efficient and effective beef-production processes. The main contributions of this architecture are (i) the ability to use everything mining to improve the knowledge of the system and decision-making processes, and (ii) three ACODAT for real-time analysis for sustainable and environmentally-friendly livestock production. The results are encouraging since the ACODAT allows smart management of the beef-production process, naturally introducing artificial-intelligence techniques to develop these tasks. Particularly, modeling using ACODAT allows an adequate description of a precision livestock process. Likewise, the preliminary results of some of the tasks of ACODAT are stimulating because they allow evaluating the feasibility of the proposal. For example, a first task for the identification of cattle fattening has a Mean Absolute Error (MAE) of 5.4 kg, which will be used by ACODAT to identify anomalies in the fattening process. The instantiation of the animal-fattening cycle shows the viability and robustness of this proposal.

1. Introduction

The *Food and Agriculture Organization* (FAO) states that sustainable development is the management and conservation of natural resources, and the orientation of technological and institutional change to ensure the continuous satisfaction of human needs for present and future generations [1]. In parallel to the growing world-population, the demand for animal protein is also increasing. Countries are reviewing the growth of animal production to meet the growing demand for animal protein [2]. On the other hand, animal-care is critical to design sustainable systems, given the loss of forest lands associated with a growing demand for meat. Particularly, environmental-sustainability issues are acute. In the context of beef cattle, (i) water use, (ii) land use, (iii) biomass appropriation (for example, the animal biomass as the feces and urine are natural fertilizers) and (iv) greenhouse-gas emissions are, for example, typically higher per unit of edible product, in beef systems, than in any other livestock systems [3].

In particular, this work considers the relationship between pasture quality and animal feed in the *Precision livestock farming* (PLF)

* Corresponding author at: CEMISID, Universidad de Los Andes, Merida, Venezuela.
E-mail addresses: rjgarciah@eafit.edu.co (R. García), aguilar@ula.ve (J. Aguilar), mtorobe@eafit.edu.co (M. Toro), perezn@ula.ve (N. Pérez), angelpinto@unisinu.edu.co (A. Pinto), elrosariocattle@gmail.com (P. Rodríguez).

https://doi.org/10.1016/j.jii.2022.100425
Received 19 August 2021; Received in revised form 3 October 2022; Accepted 7 December 2022
Available online 10 December 2022
2452-414X/

framework [4]. Thus, this work focuses on PLF but using existing technologies in precision agriculture farming (PAF) to improve forage quality. In this way, it is necessary to consider specialized machinery for seeding and fertilizing, the use of different sensors to obtain accurate information about the terrain and weather, accurate soil analysis, among other aspects, to improve animal production, harvest forage more efficiently, and reduce environmental impact [5].

Specifically, knowing the conditions of the farm helps to define an adequate feeding for the animal [6]. A good combination of forage and soil guarantees optimal results in cattle and dairy farms. The relationship between the soils of cattle farms, the types of fodder that are planted on them, and the cattle on the ranch, has recently been used to define strategies to improve the productivity and sustainability of the cattle sector [7]. In this sense, decisions are necessary on the quality of the seed, the fertilization plan, the irrigation system, phytosanitary management, weed control, and grazing programming, among other aspects, as important factors to increase yield forage and nutritional value to improve livestock productivity.

An initiative that aims to improve sustainability is PLF through the monitoring of animals in a herd. Examples of PLF are the following works. Qiao et al. [8] identified livestock by video, using artificial intelligence techniques such as Neural Networks. In addition, Denis et al. [9] used regularized change-point-estimation and the k-means algorithm for a better understanding of the lactation process. Likewise, Achour et al. [10] applied unsupervised automated-monitoring of dairy cows' conduct based on an inertial magnitude unit attached to their reverses. Finally, Guo et al. [11] identified the animal poses based on bilateral symmetry applied to the measurement of cattle body in point clouds. All these works have been focused on improving production by means of PLF. In this sense, PLF is a new approach – based on the use and development of information and communication technologies – to achieve automatic, accurate, real-time monitoring and analysis of animal behavior, which helps the producer to make decisions [12].

Previous works on PLF have not taken into account *soil variables*, which, together with *animal activity*, could lead to create robust models to improve grazing methods, and, as a consequence, to a higher production [12]. The objective of this work is to combine grazing with animal health using the paradigm of *autonomous cycles* [13]. This paradigm defines a set of data-analysis tasks for the self-management of a system. *Autonomous cycles of data analysis tasks* (ACODAT) provide cattle ranchers with tools for decision-making. Examples of the benefits of the use of ACODAT, in PLF, could be: (i) to maintain a high production of good-quality forage for the longest period of time, (ii) to maintain a favorable balance among forage species, (iii) to obtain efficient use of the forage produced, (iv) to achieve a profitable livestock production, and (v) to prevent diseases.

Each of the above benefits could use the specification of an ACODAT. In addition, to achieve these benefits, the beef-production farms must have the technology to capture and process the necessary data to build the knowledge models defined in an ACODAT. The main contributions of this research are the following.

- The use of the autonomous-computing paradigm to define an architecture for the self-managing of beef-production farms.
- The ability to integrate various mining techniques (e.g., data mining and process mining) to generate knowledge models of the farm.
- The definition of several ACODAT – with different goals about the beef-production process – to allow real-time analysis and to help decision-making for sustainable and environmentally-friendly livestock production.
- The integration of multiple variables about soil, animal welfare, pasture, and climate for the analyses.

This is a theoretical proposal that aims to show the feasibility of using the ACODAT concept to model PLF. The preliminary results of some of the analysis tasks that compose the ACODATs are encouraging. For example: the ACODAT that supervises the animal fattening process is composed of a task that defines a cattle weight identification model using machine learning techniques for the detection of anomalies [14], and a system to diagnose the anomaly in the cattle fattening process in rotational grazing using a fuzzy classification system [15]

This article is organized as follows. Section 2 presents related work. Section 3 introduces ACODAT and shows the methodology used in this work. Section 4 shows the proposed architecture for beef production in the context of PLF and, briefly, describes the three *autonomic cycles* (AC) for the self-management of a beef-production process. After, Section 5 describes the specification of an AC for self-planning of paddock rotation and its instantiation in a case study. Finally, Section 6 shows a comparison with previous work, and the article ends with conclusions in Section 7.

2. Related works

With PLF and data-mining tasks, considerable progress has been made in the use of tools to – routinely – cover and collect information from animals and farmsteads in a lower laborious manner than before, generating large volumes of data [16]. In what follows, this section presents related works on PLF and grazing with animal health.

For the problem of grazing automation and animal health, Segerkvist et al. [17] proposed a method, based on an unmanned automatic precision-weighing system – that can be used in pastures –, which alert farmers when animals show abnormal weight-gain curves. This work is – primarily – focused on detecting pasture-borne nematode-parasite infections that reduce calf weight-gain.

For the problem of weighing, Feng et al. [18] proposed a dynamic-weighing algorithm, for cows, based on support vector machines and empirical wavelet transforms for classification. The dynamic-weight curve is obtained using a weighing device placed along a cow travel corridor, and data is preprocessed using signal acquisition, feature extraction and normalization techniques. The results are divided into three levels during the movement: low, medium and high. Finally, recently, a model of normal weight identification in cattle was designed [14].

In addition to automatic weighing, monitoring is also important — as many operations are performed manually in livestock farms. Jung et al. [19] proposed a livestock-tracking system based on *Wireless Sensor Networks*. The livestock-tracking system can monitor farm animals using the *Internet of Things* (IoT) and cloud platforms. Through a collar on the neck of an animal, using IoT equipment, the system monitors the activity of the livestock. Fuentes et al. [20] proposed an artificial-intelligence system to increase milk quality by reducing heat stress. Particularly, rising global temperatures and climatic anomalies, such as heat waves, are affecting heat-stress levels of farm animals. These impacts have detrimental effects on the milk quality and productivity of dairy cows. Fuentes et al. argued that their system allows to – automatically – assess animal welfare, productivity and milk quality.

A task in monitoring is tracking. To solve this problem, Vayssade et al. [21] proposed a method to process images, taken by a commercial drone, to automate the tracking of animal activities. Their method, automatically, detects goats from images and tracks their activity using a combination of thresholding and supervised-classification methods. In another work, using drones, Li et al. [22] deployed a cluster of Unmanned Aerial Vehicles (UAV) to – autonomously – track and monitor livestock – such as cattle and sheep – in a pasture. Li et al.'s goal was to find the optimal deployment of UAV to minimize the average UAV-Animal distance, using a standard k-means clustering algorithm.

In another work using images, Benze et al. [23] designed and implemented a computer-vision system for cow individual food-intake measurement, based on deep Convolutional Neural Networks, and a low-cost RGB-D (Red, Green, Blue, Depth) camera. Timmerman et al. [24] designed a multi-layer monitoring support system to help the poultry farmer. The multi-layer support system consists of a static system for flock observation (an existing PLF technology), and robot(s) to observe bird health and behavior, in order to perform daily routine tasks.

Finally, Germani et al. [25] designed and implemented an IoT architecture to continuously monitor livestock, in barns, during grazing. Germani et al. adopted the *LOng RAnge* (LoRa) *low-power wide-area network* (LPWAN) technology to cover diverse environments, and a suitable configuration of web services to perform data storage, analysis and visualization.

All the previous works above focused on monitoring or capturing animal behavior; however, in the case of grazing, they did not use soil data. Similarly, they did not combine animal welfare with pasture intake. These limitations were also highlighted in a recent systematic literature review [12]. To overcome such limitations found in previous works, this article focuses on the use of soil variables for optimal decision-making on grazing and autonomous beef production.

3. Autonomous cycles of data analysis tasks

This section presents a background on data-mining tasks in PLF and ACODAT.

3.1. Definition of ACODAT

Autonomous computing is a paradigm in which the computing system as a whole offers much more capabilities than the sum of its parts, with self-management capabilities to adapt to its environment [26]. *Autonomic cycles of data analysis tasks* (ACODAT) have been defined by Aguilar et al. [27–29] as a set of data analysis tasks that act autonomously to supervise and/or control a process. These Data Analysis tasks are based on knowledge models (e.g., prediction, description, diagnosis, among others), and they interact and interrelate with each other according to the objectives of the cycle. Each data analysis task has a different function: (i) to observe the process, (ii) to analyze and interpret what is happening in the process, and (iii) to make decisions to improve the process.

ACODAT has been used in different domains; for instance, in smart classrooms [27,29] and Industry 4.0 [30–32]. These works based on ACODAT present a different way of introducing the autonomous computing paradigm in smart classrooms and in industry 4.0 to previous works [33–35], since ACODAT is based on data analysis tasks that can be self-managed.

ACODATs are designed using the *Methodology for the development of data-mining applications based on organizational analysis* (MIDANO) [36]. MIDANO integrates data analysis tasks into a closed-loop that can solve complex problems. In this sense, it is essential to integrate data-analysis tasks in a consistent way to generate strategic knowledge useful to achieve business objectives. A detailed description of the function, of each task category, in ACODAT, is explained in what follows.

- ***Monitoring:*** Tasks responsible to observe the monitored system. These tasks must capture data and information about the behavior of the system. In addition, these tasks are responsible for data preparation (e.g., preprocessing and selection of relevant features) for the following steps.
- ***Analysis:*** These tasks interpret, understand and diagnose what is happening in the monitored system. Particularly, these tasks allow building *knowledge models* from the dynamics observed in the system, oriented to know what is happening in the system.
- ***Decision-making:*** These tasks define and implement the necessary actions, based on previous analysis, to improve or correct the failures of the monitored system. Thus, these tasks affect the dynamics of the monitored system. The effects of the decision-making tasks are evaluated by the monitoring and analysis tasks, restarting a new iteration of the cycle.

The development of these specialized tasks (Monitoring, Analysis, and Decision-Making) depends on the context. For example, if one wishes to develop data-based models for diagnosing or predicting behavior, one can use the "Cross-Industry Standard Process for Data Mining" (CRISP-DM) methodology. On the other hand, if one wishes to develop a decision-making model using expert knowledge, one can use the "Knowledge Discovery in Databases" (KDD) methodology. In Section 6.2, we present a discussion about it.

In this work, we will propose a PLF approach based on ACODAT. It is the first time that ACODAT has been applied to PLF, in a deployment environment as broad and complex as farms with cattle rotation. In addition, it is the first time that ACODAT has been proposed in a context with a great diversity of data sources (see Section 4.2), which has a large number of sensors to be considered.

3.2. Methodology to develop ACODAT

The methodology used in this work is MIDANO [36]. MIDANO is used to develop data-mining applications based on organizational analysis. MIDANO is designed to develop applications based on ACODAT. MIDANO consists of three phases.

Phase 1. Identification of the sources of knowledge in an organization: The main objective of this phase is to know the organization, its processes and its experts, in order to define the objective of the application of data-analysis techniques in the organization. In addition, in this phase, a specification of the ACODAT to be developed is made.

Phase 2. Data preparation and processing: This process is based on the extraction of data from its sources, transformation of the data, and the load of the data into the AC data warehouse. To carry out this process, a feature engineering process is carried out in order to select the main variables of the studied process. Finally, a mineable view is created, which is composed of the description of all the variables of interest.

Phase 3. Development of the ACODAT: This phase aims to implement the different data analysis tasks of the ACODAT, which generate the required knowledge models (e.g., predictive models and descriptive models). This phase ends with the implementation of a prototype of each AC. During this phase, experiments are performed to validate the knowledge models generated by the data-analysis tasks.

In this research, MIDANO has been used for the definition of ACODAT to improve the production process of a beef farm. The data sources analyzed were animal-weighing history, animal welfare, soil quality of each paddock, climate, temperature, pastures, paddock rotations, and sale and purchase price of animals. Table 1 summarizes the use of MIDANO in this work.

4. Proposed architecture

This section presents the proposed architecture for beef production based on PLF. This is one of the main contributions of this work.

4.1. General architecture

In this research, we propose an architecture for beef production in PLF (see Fig. 1), based on the Autonomic-Computing Paradigm [26,37]. This paradigm is an essential element that guarantees the autonomy and adaptability of the production process. Autonomic properties, like (i) self-configuration, (ii) self-planning, and (iii) self-optimizing, are developed to endow autonomy in the production process. This architecture has four layers: (i) monitoring, (ii) network, (iii) data processing and (iv) business; these layers are described below.

1. ***Monitoring:*** This layer captures soil, water, animal behavior and climate variables through IoT-based sensors. After, data is stored in a cloud server. This layer can be extended by allowing the effective integration of new types of sensors [38]. Some examples

Table 1
Use of MIDANO phases in this work.

Phases	Use
Phase 1	Analysis of grazing and animal-health management process. For that, this research proposes several ACODATs to improve the beef-production process.
Phase 2	Identification of data sources (e.g., pasture, soil, animal weight) and definition of the multi-dimensional data model
Phase 3	Implementation of the ACODAT for the beef-production process. This paper presents the instantiation of one of them.

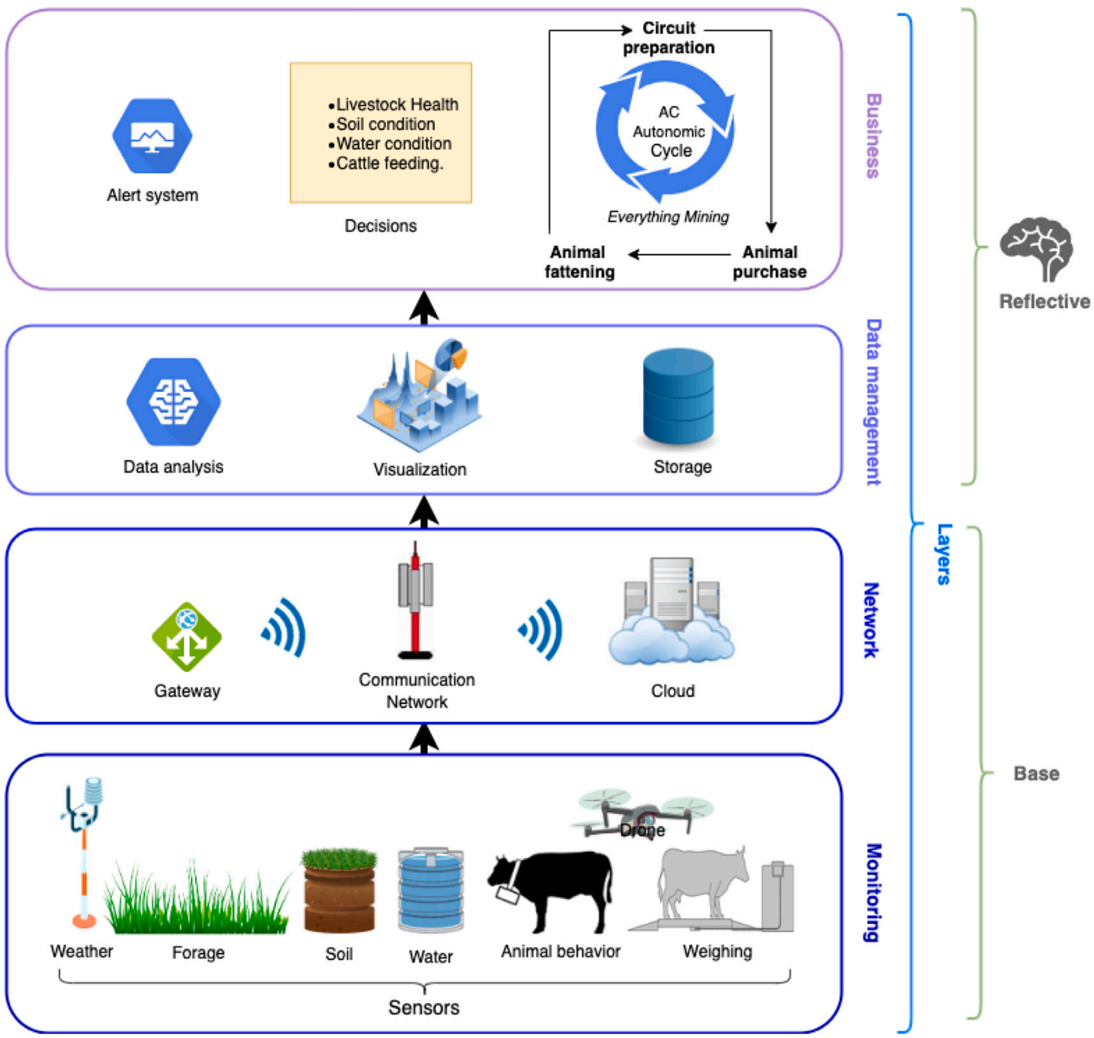


Fig. 1. Autonomic architecture for a beef-production process based on PLF.

of variables captured directly by the sensors, or derived from the sensed data, are the amount of forage per paddock, amount of quality forage per paddock, amount of water in the drinkers, amount of rain per day, movement of animals, pasture yield, among others.

2. ***Network:*** This layer is in charge of communication between the sensors and the server in the cloud.
3. ***Data management:*** This layer stores, verifies, pre-processes, and protects data. These data will be made available to the beef farmer.
4. ***Business:*** This layer is in charge to improve the beef-production process using ACODAT for decision-making. Particularly, data analysis tasks based on everything-mining techniques generate knowledge for the automation of the beef-production process (see the Reflexive Layers in Fig. 1) [30]. Thus, the analytical tasks generate useful information. Some examples of decisions or information that will be obtained in this layer are: the optimal stocking capacity per paddock supported by the livestock farm, detection of animal diseases, forage status per paddock, animal welfare, and weed detection. In addition, the architecture has an alert system – in the reflexive layer – used autonomously by the tasks that require alerts.

The next sections explain in detail the ACODATs we are proposing for holistic management of the entire beef production process, and a case study to explain in detail the implementation of the data analysis tasks for one of them.

4.2. ACODAT in beef production

This section deals with AC design for the beef-production process. We propose a system of ACs (named ACPLF-000, see Fig. 2) composed of three ACODAT: (i) preparation of the circuits, (ii) animal fattening, and (iii) animal purchases. ACPLF-000 allows the (i) self-management, (ii) self-planning, and (iii) self-supervision of the production process.

The goal of ACPLF-000 is the self-planning of the rotational grazing in the process of beef production. Specifically, the objective of each AC of ACPLF-000 is the following.

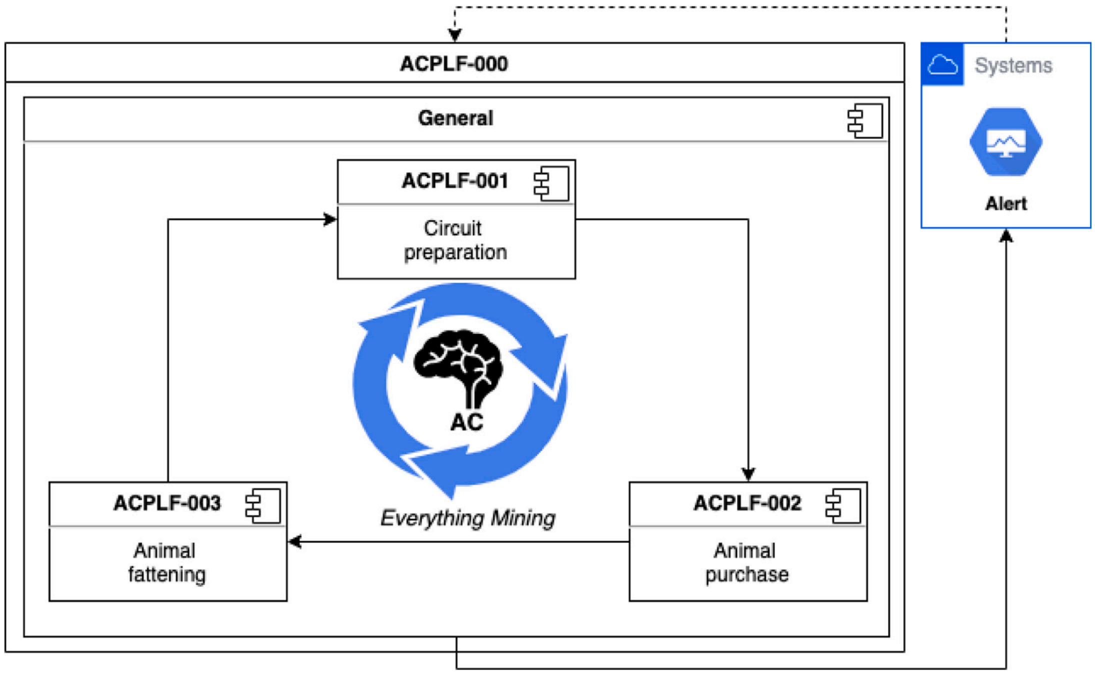


Fig. 2. ACPLF-000: AC design for the beef-production process.

Table 2
Description of the tasks of ACPLF-001. Abbreviations: DB = Data Base.

Task name	Knowledge models	Data sources
1. Determine the best seed for a specific paddock.	Recommendation model	Farm DB (Sensor data), Seed-sale website
2. Determine the amount of weeds	Detection model	Farm DB (Drone Data)
3. Determine the fodder for each paddock.	Detection model	Farm DB (Drone Data)
4. Determine how many animals can be in the pasture depending on the forage (Carrying capacity for paddock).	Estimation model	Farm DB (Sensor data, Drone Data)

ACPLF-001 (Circuit preparation): This is in charge to prepare paddocks. As an example, to verify the state of: (i) drinking troughs, (ii) electric fences, (iii) pasture capacity, and (iv) animal-carrying capacity per circuit.

ACPLF-002 (Animal purchase): This is in charge (i) to select the best supplier (i.e., another livestock farm or livestock auction), and (ii) to select the animal lot with the best characteristics.

ACPLF-003 (Animal fattening): This is in charge to manage the animal's fattening process. This cycle carries out the tasks of (i) animal weighing, (ii) animal vaccination, (iii) definition of the forage capacity per paddock, (iv) equipment verification, and (v) rotation; depending on the planning process. As an example, if an animal reaches a weight greater than 450 kg, then this animal is moved to lots – with special pastures – for its fattening completion.

These three ACs can be executed – in parallel – when there is a production in process, and – in series – when a new production is started. As an example, when a new production is started, it is necessary (i) to prepare the land, (ii) to wait for the grass to grow, (iii) to measure the forage to calculate the stocking rate, (iv) to purchase the animals, and, finally, (v) to monitor their fattening.

4.2.1. ACPLF-001 specification: Circuit preparation

The aim of this AC is to verify and configure the process of beef production based on the objective of efficient production. Mainly, this cycle is composed of three tasks (see Fig. 3).

The first task, named the Forage task, is in charge to recommend grass seeds based on the soil characteristics of each paddock. The second task has two subtasks, the first subtask is in charge to detect weeds to determine in which paddock and with what intensity the weeds are found, and the second subtask must calculate the forage capacity per paddock. Finally, the last task estimates the animal carrying per paddock. Table 2 shows the general description of each task of this AC.

Task 1. Determine the best seed for a specific paddock: This task is a grass-seed recommendation model that uses the soil sensors of the paddocks. This task captures the physicochemical variables of the soil, and, after, it determines which is the best seed to sow depending on the need and type of paddock. Examples of the types of paddocks are (i) quarantine, (ii) fattening and (iii) finishing. In addition, in this task, the attributes and prices of each seed are reviewed on specialized seed websites.

Task 2. Determine the amount of weeds in each paddock: This task detects the amount of weeds in the paddock (e.g., using images of the paddocks).

Task 3. Determine the fodder of each paddock: This task determines the weight of grass (biomass) per paddock (e.g., using images of the paddocks). Biomass is a fundamental input to calculate animal-carrying capacity.

Task 4. Determine Carrying capacity per paddock: This task determines how many animals can be, in the pasture, depending on the forage. This task invokes an optimization system that defines how many animals each paddock can support. This system uses variables such as climate, forage capacity, pasture type, pasture quality and animal welfare (e.g., percentage of shadow, water availability and feed supplements per paddock).

ACPLF-001 configures the initial conditions for the process of animal fattening on the farm. ACPLF-001 computes initial parameters, such as the number of animals per paddock and pasture inventory.

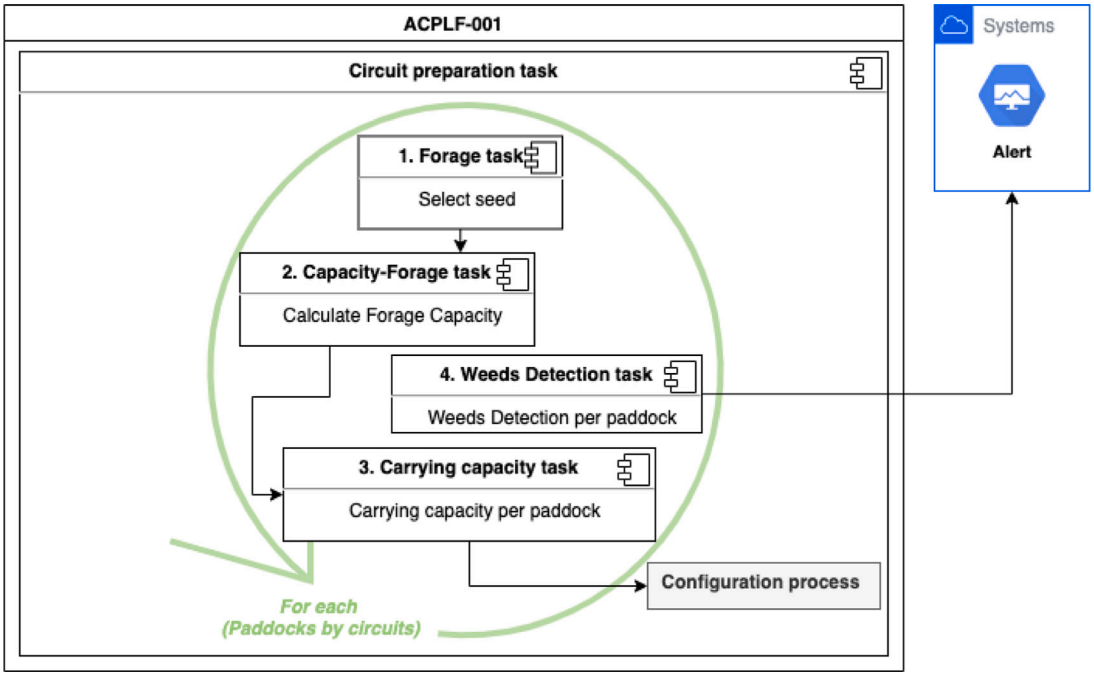


Fig. 3. ACPLF-001: Circuit preparation.

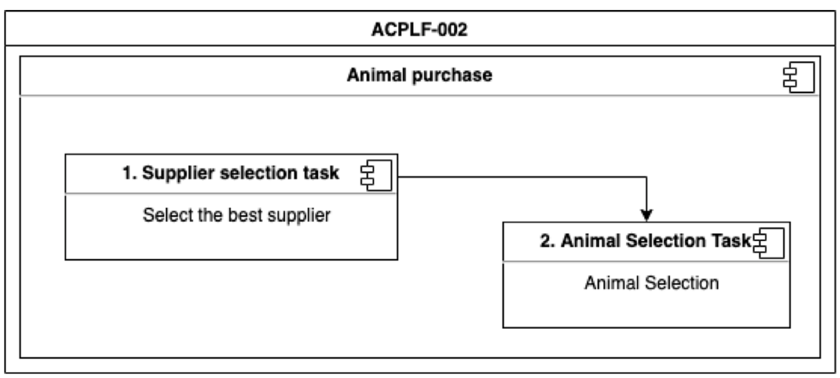


Fig. 4. ACPLF-002: Animal purchase.

Table 3
Description of the tasks of ACPLF-002. Abbreviations: DB = Data Base.

Task name	Mining techniques	Data sources
1. Determine the best supplier	Classification model	Social network, Cattle-auction website, Farm DB
2. Determine the best herd of animals	Classification model	Social network, Cattle-auction website, Cameras

4.2.2. ACPLF-002 specification: Animal purchase

The objective of this AC is the selection of suppliers and lots of cattle for purchase. This cycle is composed of two tasks (see Fig. 4). Table 3 shows the data sources used, in this AC, for each task.

Task 1. Determine the best supplier: This task ranks the best cattle supplier using different criteria such as (i) location, (ii) price, (iii) healthy-animal history and (iv) delivery time.

Task 2. Determine the best herd of animals: This task uses a model that estimates the quality of livestock using different data sources. As an example, this task can estimate (i) breed-purity percentage, (ii) best animals and (iii) health condition.

ACPLF-002 is in charge (i) to select the best animal suppliers (using the score of a supplier in the system), and (ii) to select the best herd of animals for purchase (considering season and market prices).

4.2.3. ACPLF-003 specification: Animal fattening

This AC is in charge to monitor animal's fattening by performing (i) animal weighing, (ii) animal vaccination and (iii) animal rotation (see Fig. 5). In this AC, there is a planning process of animal rotation, depending on different characteristics, such as (i) forage, (ii) climate, (iii) animal health, (iv) soil and (v) season. Table 4 shows data sources used, in this AC, for each task.

Task 1. Weighing task: This task uses the history of the previous weighing and estimates cattle weight and health status.

Task 2. Vaccination task: This task is based on a prescriptive model to determine the actions to follow in the vaccination plan (including deworming and vitamins). The main objective of this task is to reduce costs and improve profits.

Task 3. Rotation task: This task uses data from sensors located throughout the farm, and other data sources (e.g., images of the farm,

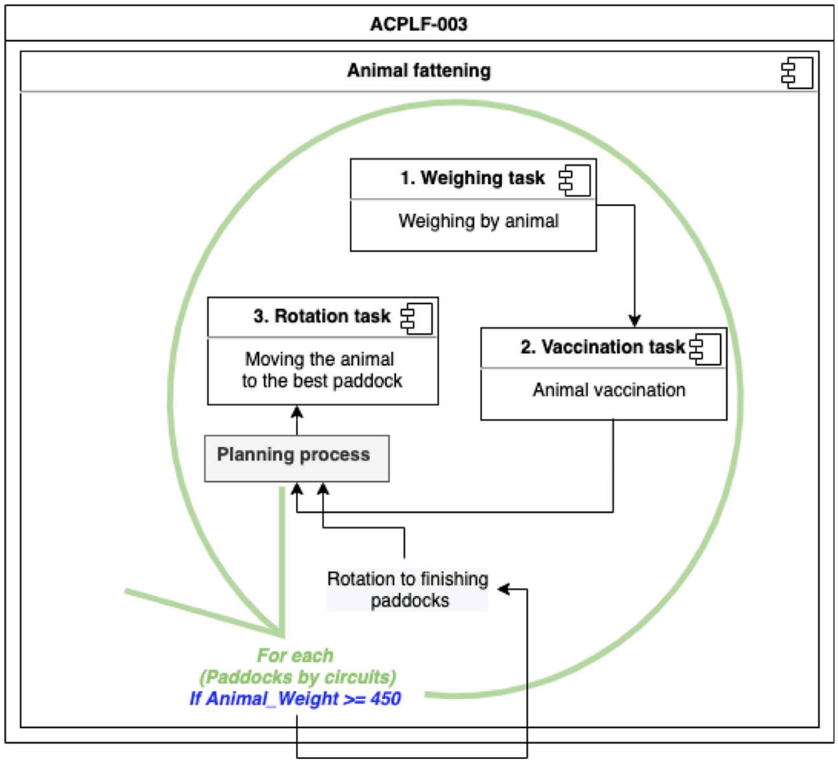


Fig. 5. ACPLF-003: Animal fattening.

Table 4
Description of the tasks of ACPLF-003. Abbreviations: DB = Data Base.

Task name	Mining techniques	Data sources
1. Weighing task	Estimation model	Farm DB
2. Vaccination task	Prescriptive models	Farm DB
3. Rotation task.	Assignment model	Farm DB

Table 5
Predictions generated by the weighting task.

Lot Id	Animal Id	Age	Weight	Predicted weight
L20-034	00732	1,5	330	327
L20-034	00733	1,6	320	319
L20-034	00734	1,6	300	325
L20-034	00735	1,4	340	338

Table 6
Prescription generated by the vaccination task.

Lot Id	Animal Id	Prescription
L20-034	00734	Anti-partisan
L20-034	00734	Genabilic acid
L20-034	00734	Mineralized salt

climate, pasture type and animal-welfare variables), to assign the best paddock available for weight gain.

4.2.4. Data model

The multidimensional data model, for the previous ACs, is defined in this section.

Fig. 6 describes the multidimensional data model required by the ACODAT for smart farm management. Each dimension represents and characterizes each information source (animal, paddock, etc.) necessary during the management process. The data model in Fig. 6 includes data from different sources, specifically, from farm databases (e.g., livestock-inventory systems), climate data obtained from the Internet, and sensor data captured in the monitoring layer that are in cloud servers, among others. Data from each source is included in a different dimension in the data model, according to its characteristics. The dimensions are the following.

Weather dimension: Contains weather data; for instance, temperature and the amount of water on the farm.

Paddock dimension: Stores data of the paddocks on the farm; for instance, area and forage capacity.

Soil dimension: Stores soil data; for instance, physicochemical variables of each paddock.

Animal dimension: Stores animal data; for instance, weight, breed and age.

Lot dimension: Stores data of the batches of animals that are in the process of fattening.

Grass dimension: Stores the data of the pasture; for instance, its type, maximum height and minimum height.

In general, data is extracted from different sources following different strategies. For example, when they are extracted from farm databases they follow a traditional data extract-transform-load approach, when they are extracted from the Internet is used a natural language processing approach, and so on for the rest of the sources. Then, they go through a phase in the data management layer that is in charge of data pre-processing, which includes data cleaning and normalization, the elimination of invalid and atypical data, among other things, and culminates with a feature engineering process that allows determining the variables required for the construction of the different knowledge models of the ACODAT data analysis tasks.

5. Case study: “El Rosario” cattle farm

For this case study, this section presents the experimental context and the instantiation of ACPLF-003 (Animal monitoring).

5.1. Experimental context

In this case study, we used data from “El Rosario” cattle farm, located 5 km from the city of Monteria, Colombia. This farm focuses

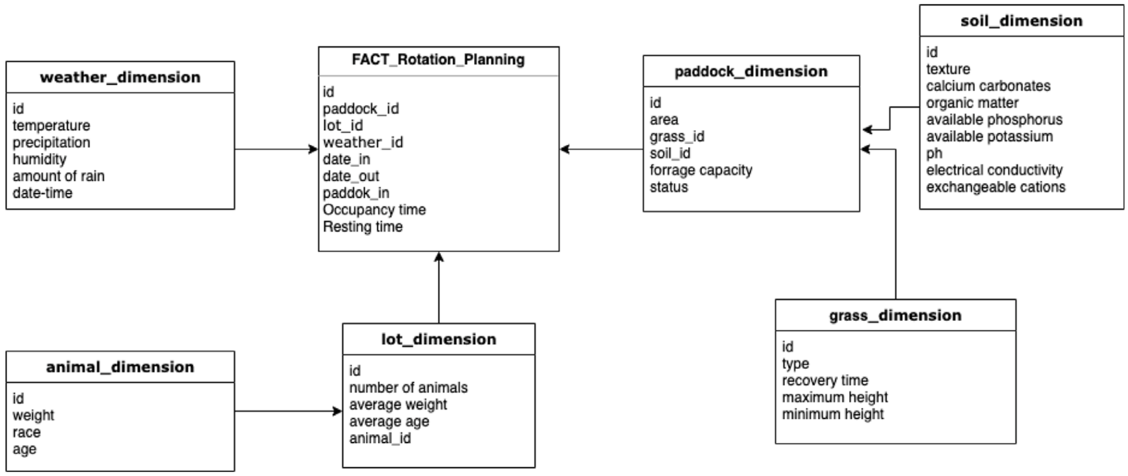


Fig. 6. Multidimensional model.

Table 7
Assignment model generated by the rotation task.

Lot ID	Quantity of animals	Average weight	Date in	Date out	Paddock	Occupancy time	Resting time
L20-034	28	300	4/01/20	8/01/20	P045	4	45
L20-034	28	303,6	8/01/20	12/01/20	P048	4	35
L20-034	28	305,4	12/01/20	14/01/20	P049	2	36
L20-034	28	307,2	14/01/20	16/01/20	P050	2	37
L20-034	28	309,9	16/01/20	19/01/20	P055	3	38
L20-034	28	311,7	19/01/20	21/01/20	P052	2	39
L20-034	28	314,4	21/01/20	24/01/20	P059	3	45
L20-034	28	316,2	24/01/20	26/01/20	P060	2	30
L20-034	28	318	26/01/20	28/01/20	P065	2	35

on beef-production with fattening cattle. The farm is certified in good cattle-raising practices by the *Colombian Agricultural Institute (in Spanish, ICA)* [39]. The farm is on the road to obtain the *Rainforest Alliance certification*. Rainforest Alliance is an international non-governmental organization that works to conserve biodiversity and ensure sustainable livelihoods [40].

To illustrate the functionality of ACPLF-000, this case study discusses ACPLF-003, in an animal monitoring process, according to the following scenario. The cattle farm has 18 lots of animals with an average of 45 animals per lot, equivalent to 810 fattening animals. These animals begin their fattening process with an average of 300 to 350 kg, with an average age of two years old. The animals must arrive to the slaughterhouse with an average of 450 to 550 kg, in a maximum of 14 months of fattening (which means they gain an average of 20 kg per month).

In the process of weight gain, the farmer must have enough food, water and nutrients to meet the weight gain demands required by the market. The fattening process requires a constant monitoring of different variables; for instance, climate, forage, animal weight and animal health. During the 14 months of fattening, the farmer, based on his experience and the advice of a veterinarian, or when necessary due to any eventuality that may arise during the process (e.g., infection, worms or skin cuts), vaccinates – frequently – the animals with antiparasitics, gives vitamins.

During the fattening, decisions are also taken about the paddocks, and the batch of animals that should be moved to based on the farmer's experience. The farmer chooses, empirically, which paddock should be occupied or not depending on (i) forage capacity, (ii) watering capacity, and (iii) which paddock has shade to improve animal welfare.

5.2. Instantiation of ACPLF-003: Animal fattening

The instantiation of ACPLF-003 must consider, for instance, disease detection, vaccination plan, paddock availability and climate. The following steps describe how ACPLF-003 is instanced for this case study.

Table 8
Comparison with previous works.

	Criterion 1	Criterion 2	Criterion 3	Criterion 4
[17]	✗	✓	✓	✗
[18]	✗	✓	✗	✗
[19]	✗	✗	✗	✗
[20]	✗	✓	✓	✗
[21]	✗	✗	✓	✗
[22]	✗	✗	✓	✗
This work	✓	✓	✓	✓

Task 1. Weighing task: The first task is to – automatically – determine the weight of cattle, for which a prediction model is used. The prediction model is built with historical data found in the farm's livestock software. The prediction model also uses variables such as (i) breed, (ii) age, (iii) sex, (iv) climate, (v) temperature and (vi) pasture quality to explain an increase or decrease in weight. In what follows, two cases of this task are presented.

- ***Case 1:*** An animal weighs 340 kg. The model estimates that it should have gained 338 kg. In this case, Task 2 would not be performed, since the animal gained the desired weight gain.
- ***Case 2:*** An animal weighs 300 kg, but the model estimated that it should have gained 325 kg, as presented in Table 5 (see Animal Id 00734). Since the conditions were favorable for weight gain (e.g., quality of pasture and climate), in this case, a warning to the farmer is generated so that he can take the appropriate decisions.

Task 2. Vaccination task: The second task uses a prescriptive model, which is built with the (i) farm's historical data, (ii) websites selling animal-feed supplements, (iii) animal medications, and a (iv) database of bovine diseases or parasites. The prescriptive model is activated depending on the weight gain/loss in the first task. As an example, in the second case, presented in Task 1, an animal weighs 300 kg and Task 1 estimated that it should weight 325 kg. This condition would invoke the prescriptive model to define a nutrition plan with vitamins or other supplements, as presented in Table 6.

Task 3. Rotation task: In the third task, the system will use an assignment model that uses environmental data to make an intelligent planning about where it is best to move a batch of animals. The assignment model takes into account (i) the quality of pasture, (ii) resting times, (iii) occupation, (iv) climate and (v) animal welfare (e.g., percentage of shade and access to water). Thus, this task will assign the best paddock to move the batch of animals to maximize meat production based on a self-planning scheme. As an example, the assignment model of Table 7 shows the planning of animal rotation (to which paddock to rotate), the duration of the occupancy, and the resting times of the paddocks.

6. General analysis

6.1. Comparison with previous works

In this section, we propose criteria to analyze beef-production automation in the context of PLF. The criteria that have been considered are desired aspects in PLF. For example, carry out everything-mining to exploit the available data. Also, automate the entire beef production process, considering the variables that directly or indirectly affect the animal, such as the pasture. On the other hand, the third criterion points to a holistic vision of the problem that must consider the relationship between soil, forage and animal. Finally, the possibility of adding aspects of environmental sustainability to model the beef production process, in order to assess its benefits, is a very important element to consider. These general criteria are within the scope of the PLF and are aspired to be achieved. In this section, we have focused on determining whether the works comply with some of these aspects, and not on how they methodologically do so. Next, we make a comparison of this work with previous works based on these criteria.

Criterion 1: The entire beef-production process is automatized.

Criterion 2: Everything-mining techniques are used in the production process.

Criterion 3: Grazing and animal welfare are analyzed together.

Criterion 4: Efficient and environmentally-friendly production is considered.

In Table 8, a qualitative comparison with related works is made, based on previous criteria (✓means the work meets that criteria and ✗means it does not).

As shown in Table 8, related papers did not satisfy all the criteria. Specifically, in criterion 1, this research allows, by means of ACs, the automation of the whole production process. For this automation, paradigms such as multi-agent systems need to be used, in conjunction with the ACODAT architecture, to model the entire production process [41].

For criterion 2, Segerkvist et al. [17] and Feng et al. [18] worked on the estimation of animal weight based on data mining. The basis of our proposal is an autonomous decision-making through data, with knowledge extracted from animal production. Thus, this work is based on everything-mining techniques.

For criterion 3, Segerkvist et al. focused on the detection of infections that reduce weight gain, something fundamental for animal welfare. In addition, Fuentes et al. [20] focused on animal welfare, specifically, in the reduction of heat stress by providing an automatic evaluation of animal welfare. Jung et al. [19] used wireless sensor networks, to monitor farm animals, using IoT equipment and cloud platforms. Vayssade et al. [21] and Li et al. [22] worked with drones for farm-data capture and monitoring tasks. Vayssade et al. used image processing for animal activity detection, and Li et al. used drones to autonomously track and monitor livestock. This proposal takes into account animal-welfare, soil, and pasture to perform rotations, in addition, this proposal includes a prescriptive model to improve animal health.

Finally, for criterion 4, this proposal meets the criterion of efficient and environmentally-friendly production because the farmer can know what is the maximum number of animals per season that can support her/his farm. This reduces greenhouse gases, and makes the farm more productive and sustainable.

6.2. Discussion of preliminary results

As mentioned before, in this work, the theoretical design of several ACODATs for self-management of the beef production process has been presented. However, the development of some of the tasks defined in these ACODATs have been carried out in previous works. For example, the ACODAT responsible for the animal fattening supervision process has two tasks, the detection of abnormalities in the fattening process and the diagnosis of what may be happening in the animal. Both tasks have already been carried out in previous works, with very stimulating results.

For example, the task that defines the animal fattening identification model was performed in [14], which allows the detection of anomalies in cattle weight gain over time. For the development of this task, the CRISP-DM methodology was used. In that work, the performance of various machine learning techniques was compared, and an outlier detection process was performed to identify outlier weights. In general, the results showed a performance with an average Mean Absolute Error (MAE) of 5.4 kg, which is quite good for cattle weights of the order of 400 kg.

Regarding the task to diagnose the anomaly detected in the fattening process of cattle in rotational grazing, a fuzzy classification system was used in [15]. For the development of this task, the KDD methodology was used. The fuzzy classifier uses the fuzzy reasoning to determine the current situation before a given input and the genetic algorithms to optimize the rules to improve diagnosis. The results of the tests carried out indicate that the quality of the fuzzy classification system is very good. The value of the Area Under the Curve that measures the sensitivity and specificity of the system is 1 (indicates a precision of 100% in the diagnosis); and the certainty of the rules it optimizes (determines the average degree of their firing), in the end, have a value of 0.85, which is very good.

Specifically, this paper shows the implementation of the autonomous self-planning cycle as the first step toward semi-autonomous production processes, in the context of PLF. The described autonomous cycle was methodologically detailed using MIDANO, to achieve the desired results. This means that this AC achieved the designed objectives of animal-rotation planning, aiming to incorporate autonomy in the beef production process. The preliminary results indicated in this section show the viability of our proposal, and motivate the future development of all the ACODATs defined in this work.

7. Conclusions and future work directions

This paper presents an architecture to integrate and inter-operate actors in the context of PLF. The architecture combines multiple variables; for instance, soil, animal welfare, pasture, climate. The proposed architecture uses ACODAT and proposes several ACs to give autonomy to the beef-production process. In addition, this article shows the instantiation of an autonomous self-planning cycle as the first step towards efficient and effective beef-production processes.

Other results of this research are (i) the use of environmental variables to make decisions on optimal grazing, (ii) the ability to use everything-mining techniques to improve the knowledge of the system and decision-making processes (.e.g., data mining, text mining and web-mining techniques), (iii) real-time analysis for a sustainable and environmentally-friendly livestock production , and (iv) use of

PLF and PAF combined and managed by CAs capable of adapting to the beef production process. Thanks to the concept of ''autonomous data analysis cycles'', the model can self-adjust to new environmental constraints, as well as to growth/decrease in forage demand, among other events. For this purpose, several data analytics techniques are used, which autonomously work to make decisions.

Particularly, the case study shows that everything-mining techniques are needed to address self-planning in rotational grazing. This case study serves as a guide to incorporate self-management in a beef-production process. Finally, the implementation of this case study will allow us to evaluate the impact of this self-management approach of beef-production processes, in the context of PLF, based on the autonomous-computing paradigm and everything-mining techniques.

Future work is aimed to implement this framework in a simulated environment to verify the functionalities of this solution. In this sense, we plan to use historical data of ''El Rosario'' Cattle Farm to develop the different data-analysis tasks (e.g., to determine weight gains or losses in the rotation of animal flocks). Another important future work is to develop specific ACs to evaluate autonomic processes to configure and reconfigure a livestock farm, in the context of PLF, according to this architecture.

Finally, one of the biggest challenges to implement this type of system is the cost associated with the sensors for data acquisition. In addition to the optimal distribution of these sensors, Internet connectivity and stable electricity-flow on the farm are big challenges. Future studies should analyze these uncertain aspects and how to consider them in the proposed architecture.

CRediT authorship contribution statement

Rodrigo García: Conceptualization, Simulations, Formal analysis, Writing. **Jose Aguilar:** Conceptualization, Methodology, Formal analysis, Writing, Supervision. **Mauricio Toro:** Formal analysis, Writing, Supervision. **Nelson Pérez:** Writing, Funding acquisition. **Angel Pinto:** Writing, Funding acquisition. **Paul Rodríguez:** Formal analysis, Case study.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

Rodrigo García is supported by a PhD grant financed by Universidad EAFIT. All the authors would like to thank the ''Vicerrectoría de Descubrimiento *y* Creación'', of Universidad EAFIT, for their support on this research.

References

[1] L.O. Tedeschi, J.P. Muir, D.G. Riley, D.G. Fox, The role of ruminant animals in sustainable livestock intensification programs, Int. J. Sustain. Dev. World Ecol. 22 (5) (2015) 452–465.

[2] H. Erdal, G. Erdal, B. Ayyildiz, Are support policies for sustainable livestock important? Causality between animal existence and support policies: Vecm analysis for Turkey, J. Anim. Plant Sci. 31 (1) (2021) 254–264.

[3] P. Gerber, A. Mottet, C. Opio, A. Falcucci, F. Teillard, Environmental impacts of beef production: Review of challenges and perspectives for durability, Meat Sci. 109 (2015) 2–12.

[4] E. Tullo, I. Fontana, A. Diana, T. Norton, D. Berckmans, M. Guarino, Application note: Labelling, a methodology to develop reliable algorithm in PLF, Comput. Electron. Agric. 142 (2017) 424–428.

[5] B. Kumar, M. Sobhana, J. Duvvuru, C. Nikhil, G. Sridhar, Precision Agriculture Farming by Monitoring and Controlling Irrigation System Using Sensors, Vol. 101, in: Lecture Notes on Data Engineering and Communications Technologies, 2022, pp. 331–343, cited By 0.

[6] F. Sigcha, Y. Pallavicini, M. Camino, C. Martínez-Ruiz, Effects of short-term grazing exclusion on vegetation and soil in early succession of a Subhumid Mediterranean reclaimed coal mine, Plant Soil 426 (1) (2018) 197–209.

[7] M.E. Ramos-Font, M. Tognetti-Barbieri, J. González-Rebollar, A.B. Robles-Cruz, Potential of wild annual legumes for mountain pasture restoration at two silvopastoral sites in southern Spain: promising species and soil-improvement techniques, Agrofor. Syst. 95 (1) (2021) 7–19.

[8] Y. Qiao, D. Su, H. Kong, S. Sukkarieh, S. Lomax, C. Clark, BiLSTM-based individual cattle identification for automated precision livestock farming, in: IEEE International Conference on Automation Science and Engineering, Vol. 2020-August, 2020, pp. 967–972.

[9] C. Denis, E. Lebarbier, C. Lévy-Leduc, O. Martin, L. Sansonnet, A novel regularized approach for functional data clustering: an application to milking kinetics in dairy goats, J. R. Stat. Soc. Ser. C. Appl. Stat. 69 (3) (2020) 623–640.

[10] B. Achour, M. Belkadi, R. Aoudjit, M. Laghrouche, Unsupervised automated monitoring of dairy cows' behavior based on Inertial Measurement Unit attached to their back, Comput. Electron. Agric. 167 (2019).

[11] H. Guo, Z. Li, Q. Ma, D. Zhu, W. Su, K. Wang, F. Marinello, A bilateral symmetry based pose normalization framework applied to livestock body measurement in point clouds, Comput. Electron. Agric. 160 (2019) 59–70.

[12] R. García, J. Aguilar, M. Toro, A. Pinto, P. Rodríguez, A systematic literature review on the use of machine learning in precision livestock farming, Comput. Electron. Agric. 179 (2020) 105826.

[13] J. Aguilar, O. Buendia, A. Pinto, J. Gutiérrez, Social learning analytics for determining learning styles in a smart classroom, Interactive Learning Environments 30 (2) (2022) 245–261.

[14] R. García, J. Aguilar, M. Toro, M. Jiménez, Weight-identification model of cattle using machine-learning techniques for anomaly detection, in: 2021 IEEE Symposium Series on Computational Intelligence (SSCI), 2021, pp. 01–07.

[15] C. Benitez, R. García, J. Aguilar, M. Jiménez, H. Robles, Supervision System of the Fattening Process of Cattle in Rotational Grazing using Fuzzy Classification, in: 2022 XLVIII Latin American Computer Conference (CLEI), 2022, pp. 1–10, http://dx.doi.org/10.1109/CLEI56649.2022.9959950.

[16] G. Morota, R.V. Ventura, F.F. Silva, M. Koyama, S.C. Fernando, Big data analytics and precision animal agriculture symposium: Machine learning and data mining advance predictive big data analysis in precision animal agriculture1, J. Anim. Sci. 96 (4) (2018) 1540–1550.

[17] K.A. Segerkvist, J. Höglund, H. Österlund, C. Wik, N. Högberg, A. Hessle, Automatic weighing as an animal health monitoring tool on pasture, Livest. Sci. 240 (2020) 104157.

[18] N. Feng, X. Kang, H. Han, G. Liu, Y. Zhang, S. Mei, Research on a dynamic algorithm for cow weighing based on an SVM and empirical wavelet transform, Sensors 20 (18) (2020).

[19] E.Y.P. Jung Kyu Park, Animal monitoring scheme in smart farm using cloud-based system, Thaijo 15 (1) (2021).

[20] S. Fuentes, C. Gonzalez Viejo, B. Cullen, E. Tongson, S.S. Chauhan, F.R. Dunshea, Artificial intelligence applied to a robotic dairy farm to model milk productivity and quality based on cow data and daily environmental parameters, Sensors 20 (10) (2020).

[21] J.-A. Vayssade, R. Arquet, M. Bonneau, Automatic activity tracking of goats using drone camera, Comput. Electron. Agric. 162 (2019) 767–772.

[22] X. Li, L. Xing, Use of unmanned aerial vehicles for livestock monitoring based on streaming K-means clustering this work was supported by the Australian research council, IFAC-PapersOnLine 52 (30) (2019) 324–329, 6th IFAC Conference on Sensing, Control and Automation Technologies for Agriculture AGRICONTROL 2019.

[23] R. Bezen, Y. Edan, I. Halachmi, Computer vision system for measuring individual cow feed intake using RGB-D camera and deep learning algorithms, Comput. Electron. Agric. 172 (2020).

[24] M. Timmerman, R. Van Emous, J. Van Riel, B. Vroegindeweij, C. Lokhorst, Market consultation for a multi-level monitoring system with robots to support poultry farmers, in: Precision Livestock Farming 2017 - Papers Presented At the 8th European Conference on Precision Livestock Farming, ECPLF 2017, 2017, pp. 542–549.

[25] L. Germani, V. Mecarelli, G. Baruffa, L. Rugini, F. Frescura, An IoT architecture for continuous livestock monitoring using lora LPWAN, Electr. (Switzerland) 8 (12) (2019).

[26] P.J. Horn, Autonomic computing: Ibm's perspective on the state of information technology, 2001.

[27] J. Aguilar, M. Sánchez, J. Cordero, P. Valdiviezo-Díaz, L. Barba-Guamán, L. Chamba-Eras, Learning analytics tasks as services in smart classrooms, Univ. Access Inf. Soc. 17 (4) (2018) 693–709.

[28] M. Sánchez, J. Aguilar, J. Cordero, P. Valdiviezo-Díaz, L. Barba-Guamán, L. Chamba-Eras, Cloud computing in smart educational environments: Application in learning analytics as service, in: A. Rocha, A.M. Correia, H. Adeli, L.P. Reis, M. Mendonça Teixeira (Eds.), New Advances in Information Systems and Technologies, 2016, pp. 993–1002.

[29] J. Aguilar, J. Cordero, O. Buendía, Specification of the autonomic cycles of learning analytic tasks for a smart classroom, J. Educ. Comput. Res. 56 (6) (2018) 866–891.

[30] M. Sanchez, E. Exposito, J. Aguilar, Autonomic computing in manufacturing process coordination in industry 4.0 context, J. Ind. Inf. Integr. 19 (2020) 100159.

[31] J. Terán, J. Aguilar, M. Cerrada, Integration in industrial automation based on multi-agent systems using cultural algorithms for optimizing the coordination mechanisms, Computers in Industry 91 (2017) 11–23.

[32] M. Sánchez, E. Exposito, J. Aguilar, Implementing self-* autonomic properties in self-coordinated manufacturing processes for the Industry 4.0 context, Comput. Ind. 121 (2020).

[33] S. Bachir, L. Gallon, A. Abenia, P. Aniorte, E. Exposito, Towards autonomic educational cyber physical systems, in: 2019 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), 2019, pp. 1198–1204.

[34] S. Maksuti, M. Tauber, J. Delsing, Generic autonomic management as a service in a SOA-based framework for industry 4.0, in: IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society, Vol. 1, 2019, pp. 5480–5485.

[35] N. Sharma, M. Shamkuwar, P. Ramdasi, Digital dimensions of industry 4.0: Opportunities for autonomic computing and applications, in: T. Choudhury, B.K. Dewangan, R. Tomar, B.K. Singh, T.T. Toe, N.G. Nhu (Eds.), Autonomic Computing in Cloud Resource Management in Industry 4.0, Springer International Publishing, Cham, 2021, pp. 347–383.

[36] F. Pacheco, C. Rangel, J. Aguilar, M. Cerrada, J. Altamiranda, Methodological framework for data processing based on the Data Science paradigm, in: 2014 XL Latin American Computing Conference (CLEI), 2014, pp. 1–12.

[37] J. Vizcarrondo, J. Aguilar, E. Exposito, A. Subias, MAPE-K as a service-oriented architecture, IEEE Latin Am. Trans. 15 (6) (2017) 1163–1175.

[38] Y. Han, J. Ren, Q. Zhu, D. Barclay, J. Windmill, IoT and cloud enabled evidence-based smart decision-making platform for precision livestock farming, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 11691 LNAI, 2020, pp. 570–582.

[39] L. Mestra, M. Santana, D. Rios, L. Mejia, C. Ortiz, S. Paternina, Characterization of sheep feeding systems in the department of Córdoba, Colombia, Arch. Zootecnia (2020).

[40] A. Annunziata, A. Mariani, R. Vecchio, Effectiveness of sustainability labels in guiding food choices: Analysis of visibility and understanding among young adults, Sustain. Product. Consum. (2019).

[41] J. Aguilar, M. Cerrada, F. Hidrobo, A methodology to specify multiagent systems, in: N. Nguyen, A. Grzech, R.J. Howlett, L.C. Jain (Eds.), Agent and Multi-Agent Systems: Technologies and Applications, 2007, pp. 92–101.

Appendix C

Weight-identification model of cattle using machine-learning techniques for anomaly detection

Weight-Identification Model of Cattle Using Machine-Learning Techniques for Anomaly Detection

Rodrigo García
GIDITIC, Universidad EAFIT
Faculty of Engineering, Universidad del Sinú
Medellin, Colombia
rjgarciah@eafit.edu.co

Jose Aguilar
GIDITIC, Universidad EAFIT
CEMISID, Universidad de Los Andes
Universidad de Alcalá, Dept. Automática
Colombia, Venezuela, España
aguilar@ula.ve

Mauricio Toro
GIDITIC, Universidad EAFIT
Medellin, Colombia
mtorobe@eafit.edu.co

Marvin Jiménez
Dept. Faculty of Engineering, Universidad del Sinú
Montería, Colombia
marvinjimenez@unisinu.edu.co

2021 IEEE Symposium Series on Computational Intelligence (SSCI) | 978-1-7281-9048-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/SSCI50451.2021.9659840

***Abstract*—Cattle raising is an important economic activity, where livestock entrepreneurs keep track of their production and investment costs, to measure production and business profitability, based on cattle weighing. However, it is complicated for the farmer to detect if the animals they are weighing have gained the right weight. This paper proposes a framework to identify the fattening process, which can be used to detect anomalies in cattle weight-gain over time. This framework used records of animals raised and fattened at "El Rosario" farm, located at the municipality of Montería (Córdoba-Colombia), to identify the fattening process. The performance of four machine-learning techniques to identify the ideal weight from real data was compared. The algorithms used were Decision Tree (DT), Gradient Boosting (GB), regression based on K-Nearest Neighbors (KNN), and Random Forest (RF). In addition, an outlier-detection process was performed to identify anomalous weights. In general, the results showed that the DT model was the one with the best performance with an average Mean Absolute Error (MAE) of 5.4 kg.**

***Index Terms*—Identification system, Machine learning, Precision Livestock Farming, Anomaly detection**

I. Introduction

The goal of anomaly detection is to detect abnormal or atypical behavior within a system. Weight and weight gain are two of the most frequently analyzed traits in cattle [1]. Knowledge of the weight of cattle is considered of great importance in (i) growth-evaluation processes, (ii) feeding of different categories of animals at different times of the year, (iii) formation of homogeneous groups according to weight and/or size, (iv) in the use of available food resources, and (v) in observation tasks. In general, the farmer detects which of her/his animals are gaining and/or losing weight based on the experience she/he has with the breed and its performance on the farm.

Rodrigo García is supported by a PhD grant financed by Universidad EAFIT

Machine-learning techniques have been used in cattle breeding for early detection of lameness, and efficient identification of which heifers will be able to breed successfully. Machine Learning has also been used to help farmers to properly allocate resources on their farms and increase production efficiency [2].

Identification techniques propose an approximated model of a real system, based on linguistic or mathematical expressions, or an algorithm [3]. System Identification is quite a mature area that has had an interesting and productive development [4], but many problems remain. One of these problems is the definition of simple Models for Control Tuning and Performance Monitoring [4]. In this paper, it is the farmer's decision to improve paddocks and rotations or to detect animals with anomalies based on cattle weights (animals that did not gain the weight they should have gained).

This paper presents an automated identification-system approach based on regression algorithms with the predictors of gender, age and animal breed. Particularly, the response variable is the weight of the animal, describing the pasture-based fattening process in real-time. This approach is used to detect anomalies and could be a valuable tool for decision making in beef or milk production processes, for example.

This work is organized as follows. Section 2 introduces the Theoretical Framework used in this work. Section 3 shows an approach to identify whether the animals are sick or there are low-productivity paddocks on the beef-production process. After, Section 4 describes the experiments. Finally, Section 5 presents the conclusions and future works related to the obtained results.

II. RELATED WORKS

For PLF, considerable progress has been made in the use of tools to routinely monitor and collect information from animals and farms, in a less laborious manner than before, generating large volumes of data [5]. In what follows, this section presents the related works on PLF for the detection of anomalies in the production process.

A. Anomaly detection in animal welfare

Sai *et al.* [6] collected the environmental temperature, humidity, illumination and body-surface temperature of cattle in a barn, to develop an artificial-intelligence model that estimates the deep-body temperature of cattle without contact. Thus, the estimation reached the detection of anomalies, in cattle health, by collecting data on the actual environment of the cattle.

Haladjian *et al.* [7] detected abnormalities in a cow's gait pattern as a possible indicator of lameness. The detection is carried out by means of a portable motion sensor placed on the cow's left hind leg. They built an individual model of a cow's usual gait pattern during the first few minutes of use and detected deviations from this model afterward. They were able to detect deviations in the cow's gait with an accuracy of 91.1%.

Wagner *et al.* [8] conducted a study on dairy cows, in which their activities were captured by an indoor tracking system and considered as time series. The state of cows (diseases, stress, no problem) was manually labeled by animal caretakers or by a sensor for ruminal pH (acidosis). They used Fourier-based method to detect anomalies in time series.

Chung *et al.* [9] proposed a data-mining solution for the detection of oestrus, using sound data of Korean native cows (Bos taurus coreanea). They extracted the mel frequency cepstrum coefficients from sound data with a feature-dimension reduction, and used a support vector machine as an early anomaly detector. Results showed that this method can be used to detect oestrus both economically (even with a cheap microphone) and accurately (over 94% accuracy).

Kramer *et al.* [10] developed a fuzzy-logic model for classification and control of lameness and mastitis, in cows, using data of the Futterkamp dairy-research farm of the Schleswig-Holstein Chamber of Agriculture. Disease alerts, by the fuzzy-logic model, were generated using the variables milk yield, dry-matter intake, dry-matter intake behavior (number of visits at the feeding trough, time spent at the feeding troughs), water intake, activity and information about preliminary diseases as input data.

B. Anomaly detection in farm equipment

Park *et al.* [11] proposed a mechanism to detect anomalies in pig house equipment through a recurrent neural network learning model using data from sensors and environmental controllers. To predict malfunctions of each equipment, when something goes wrong with the sensor, they use the difference between the predicted value and the measured value.

C. Anomaly detection in animal behavior

Cai *et al.* [12] developed a monitoring system based on passive infrared detection is proposed to analyze daily-hog activity and abnormal behaviors. Data for 90 days (day and night) were collected in a hog room to establish the model of daily activity. To find the abnormal behaviors during the night, k-means clustering was used.

Calera *et al.* [13] presented two approaches about information services based on multi-source info, being DEIMOS-1 the major satellite contributor. The first approach is based on in-depth monitoring of the crop phenology, characterized by means of the weekly Normalized Difference Vegetation Index. The second approach is focused on drought and other anomaly detection in crops and pastures. For pastures, a specific module has been designed to detect drought occurrence by comparing current Normalized Difference Vegetation Index values with historical ones. Besides, they detected maximum livestock stocking rate and need of supplementary feeding and overstocking risk.

In this work, we detect anomalies in weight gained, using an identification model of the cattle fattening process based on rotational grazing. The identification model is based on regression techniques that determine the weight that cattle should have been gained after a rotation. In case this value is higher or lower than the actual weight of the animal, it is marked in the system as an anomaly.

III. THEORETICAL FRAMEWORK

A. Precision Livestock Farming

Precision Livestock Farming (PLF) uses biometric sensors, artificial intelligence, big data and blockchain technologies. Biometric sensors include invasive or non-invasive sensors that monitor an individual animal's health and behavior in real-time, allowing producers to integrate this data for population-level analysis [14]. Real-time information from biometric sensors is processed and integrated by big-data analytics systems that rely on machine learning and statistical algorithms to classify large data sets to provide producers with relevant trending patterns and decision-making tools [14].

The main objectives of PLF are: (i) to identify the most appropriate livestock feeding, (ii) to reduce environmental impact through efficient resource management, (iii) to manage crop processes to make a perfect synergy with livestock feeding, (iv) to ensure food safety through traceability (documentary record from production to consumption) of products, and (v) to improve animal health and crop efficiency [2].

One of the challenges of Precision Livestock Farming is to use different machine learning models to automate production processes. This implies seeking correct modeling strategies (e.g., using multi-agent systems [15], [16]) and integration of machine learning models (e.g., using the concept of autonomous cycles of data analysis tasks [17]).

B. System identification

The distinguishing features– as well as the bulk of efforts in System Identification– can, somewhat simplistically, be described as to define models $\hat{y}(t|\Theta)$ suitable to describe linear and nonlinear dynamic systems [4]. This can be summarized as a black box, where we can only observe the change of the output due to the change of the input [18], [19]. The goal of an identification system is to find a simple description of the observed behavior. If we have not some idea of what might be in the box, the identification task is more difficult; this means, it is harder to find a description of the data obtained. In real situations, what is usually done is to try a finite number of experiments and increase the complexity of the model more and more, until the resulting data are explained or justified.

For the definition of an identification model, there are several techniques in the literature. Fuzzy logic [3], [20], evolutionary algorithms [21], [22] and bio-inspired algorithms (e.g., artificial neural-networks and artificial immune-system [23]) approaches have been used. In this paper, a regression scheme will be used. Regression is the process of learning the relationship of input space X and output space Y by adjusting parameters of a mathematical function $f : X \rightarrow Y$, so that the error between the model forecast and the real data is minimized [24].

Often used Regression algorithms are, for instance, gradient-boosting regressor, adaboost regressor, linear regression, regression trees, among others [25]. Such models require a hyperparameter-optimization process. Usually, the alternatives for hyperparameter optimization are grid search or random search. With these approaches, a defined area of the hyperparameter space is explored. While random search examines a random combination of hyperparameters, grid search examines all parameter constellations with a fixed step size [24]. Research results have shown that the optimization result of random search is, in most cases, as good as that of grid search [26], and random search is usually faster.

C. Anomaly detection

Anomaly detection refers to the task of identifying rare observations, which differ from the general distribution of the data at hand [27]. Anomaly-detection approaches have the potential of summarizing the status of a multivariate system with a singular quantitative indicator, that's is typically called Anomaly Score [28]. Anomaly detection has been successfully applied in various areas of science, such as fraud detection [29], biomedical engineering [30], the oil and gas industry [31], and monitoring of rotating machinery [28], among others.

There are different strategies for anomaly detection. In this paper, we use an isolation forest approach. Isolation forest [32] uses the concept of isolation rather than measuring distance or density to detect anomalies. The Isolation Forest exploits an area partitioning procedure: the most idea underlying the approach is that an outlier would require less iterations than an inlier to be isolated, this means that it seeks out through the partitioning procedure a neighborhood of the space where only such observation lies in [28].

The partitioning procedure employed by the Isolation Forest is achieved through the creation of iTrees~~,~~: binary trees that are the results of a random partitioning procedure obtained by splitting the info supported using their features at each iteration of the algorithm. Following the above idea of Isolation Forest, it is expected that the trail to a leaf node from the basis of an iTree is going to be shorter for outliers than for inliers. Thus, the anomaly score is going to be associated with this path length: the shorter the more anomalous the point is.

D. Machine learning techniques

Machine learning can have a multitude of applications and uses. Among the many existing machine learning techniques, we will mention those used in our proposal.

Regression DT are trees whose leaves predict a real number and not a class. In the case of regression, it looks for splits that minimize the squared error of prediction squared error (the least squared deviation). The prediction at each leaf is based on the weighted mean of the node [33].

GB is a state-of-the-art prediction technique that sequentially produces a model in the form of linear combinations of elementary predictors typically DT by solving an infinite-dimensional convex optimization problem [34].

KNN searches the distances between a query and all examples in the data, selecting the specified number of examples (K) closest to the query, then votes for the most frequent label (in the case of classification) or averages the labels (in the case of regression) [35].

RF is aggregate a collection of random DTs. The goal is, instead of seeking to optimize one predictor "at a time" as in the case of a DT, to group a set of (not necessarily optimal) predictors. Since individual trees are randomly perturbed, the forest benefits from a wider exploration of the space of all possible predictors of the tree, which, in practice, results in better predictive performance. [36].

IV. Our approach

We present an identification model that can be used in different contexts (e.g., process control or supervision). Particularly, in this paper, the model is used for anomaly detection in a beef production process (see Fig 1). In this general architecture of anomaly detection, there are two main processes: one for the definition of the identification model of the cattle fattening process, and another for the diagnosis of anomalies in a beef production process.

The identification process carries out two main activities: feature engineering and identification model construction. The feature engineering consists of the extraction and description of the relevant characteristics; in this case, age, breed and gender were identified. The identification model was defined as a predictive model, and was built using machine-learning techniques from data collected and validated by zootechnics at the "El Rosario" farm located at Montería-Colombia.

The diagnostic process determines the ideal range of values of the cattle weights because the prediction technique incorporates stochastic processes, such that the same inputs can generate different results in a prediction. With this range, the expected or considered normal and atypical weights are determined by implementing a forestall isolation algorithm,

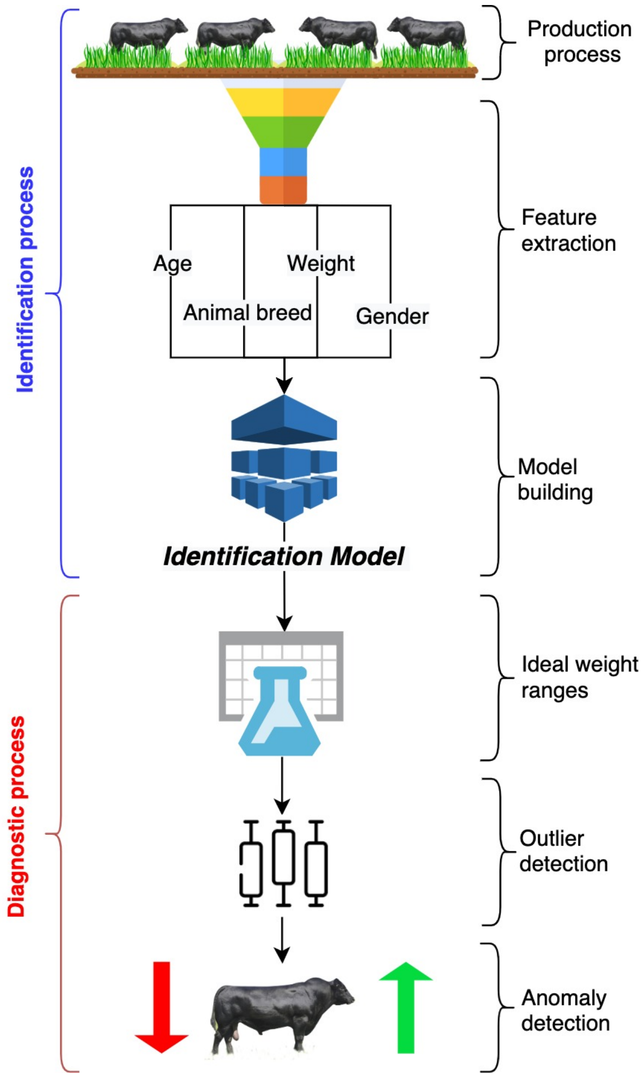


Fig. 1. Anomaly detection architecture in a beef production process.

which returns the anomaly score that is compared to the expected weight of the animal, thus detecting upward or downward anomalous weight changes.

V. EXPERIMENTS

In this section, we build a model to identify the animal fattening process by rotational grazing. We compare different regression techniques to model the ideal weight of an animal as a function of time, breed and gender. We define this identification model with a dataset provided by "El Rosario" farm, which contains the weight of different breeds of animals per lot. The dataset contained the records of 104 animals from the day of birth to slaughter, and it is evaluated by four machine learning techniques using ten fold cross-validation (80% of the records for training and 20% for testing).

A. Comparison of regression techniques

Fig. 2 shows the boxplots of the mean absolute error (MAE) of weight prediction for each technique used, which is calculated using equation 1.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \widehat{y}_i| \quad (1)$$

where N is the total number of the number of records in the evaluated dataset, y_i and $\widehat{y}_i$ are the actual weight and weight prediction by the model, respectively.

According to Fig. 2, DT and RF show very similar behaviors and the lowest MAEs with medians below 5.5 kilograms. Although GB and KNN algorithms show higher average MAEs, they do not seem to have relevant differences.

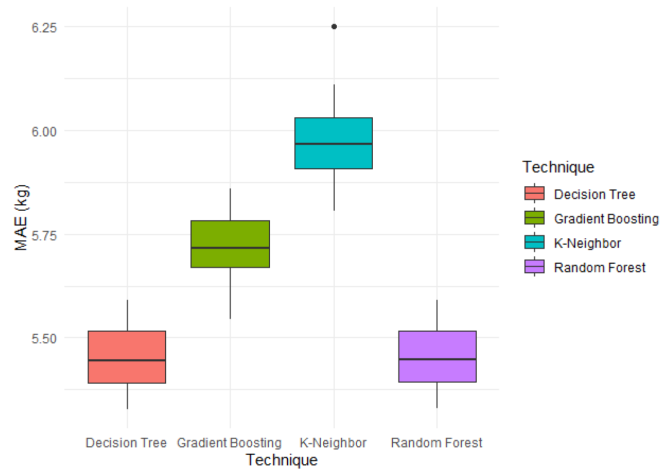


Fig. 2. MAE boxplot for each technique.

To statistically verify the existence of significant differences between the average MAEs of the predictions of the used techniques, to determine which of them is the most adequate to identify the ideal weight of the animals over timeThus, an analysis of variance (ANOVA) was carried out to test the hypothesis contrast, presented in Table I.

TABLE I
ANOVA HYPOTHESIS TEST.

Hypothesis	Description of Hypothesis
$H_0 : \mu_{DT} = \mu_{RF} = \mu_{GB} = \mu_{KN}$	The means of the MAE are the same.
$H_1 : \mu_i \neq \mu_j$, for some $i \neq j$	At least one of the means is different.

The following validates the fulfillment of the necessary assumptions in an ANOVA. On the one hand, the scatter plot of the residuals as a function of time presented in Fig 3 shows no signs of correlation between the residuals and the time order of data collection. This suggests that the assumption of independence of the errors is fulfilled. On the other hand, the empirical distribution of the residuals has a bell-shaped behavior, quite symmetrical, unimodal and very similar to the Normal distribution, suggesting that the errors follow a Normal distribution as the probability (see the histogram in Fig 4).

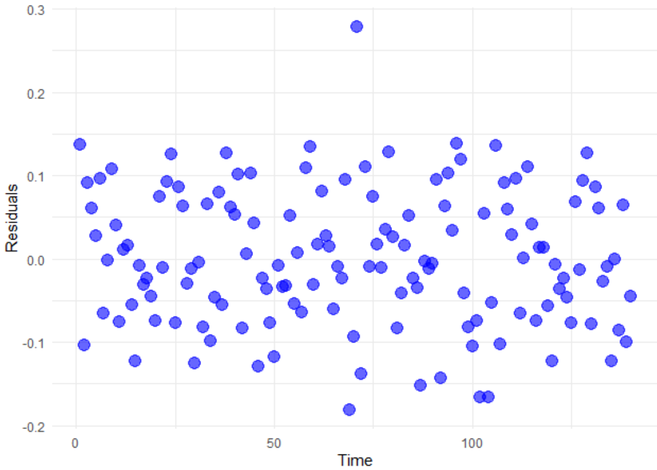


Fig. 3. Plot of residuals versus time.

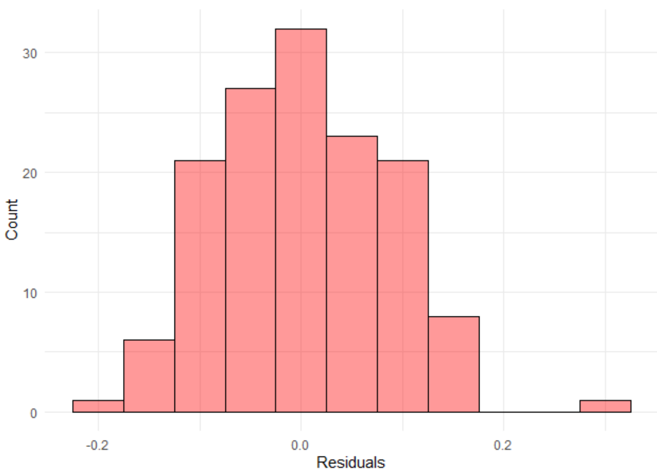


Fig. 4. Histogram of residuals.

The scatter plot of the residuals against the adjusted values, in Fig 5, shows a high homogeneity in the variability of the residuals; no relevant changes are observed in the scatter as a function of the adjusted values, indicating that the errors have a constant variance. To verify what was obtained with the graphic analysis, and setting a significance level of 0.05, the Shapiro-Wilk Normality test and Bartlett's test of homogeneity of variances were performed, which yielded p-values of 0.1011 and 0.252, respectively. These results confirm that the assumptions necessary for the validity of the analysis of variance are met.

The ANOVA results are presented in Table II. Since the p-value of the test ($< 2.2e-16$) is less than the 0.05 significance level, it is statistically verified that there are significant differences between the means of the MAEs of the prediction algorithms used. Therefore, a multiple comparisons process is performed using Fisher's least significant difference (LSD) procedure and Tukey's test with which the same results are obtained, which are shown in Table III.

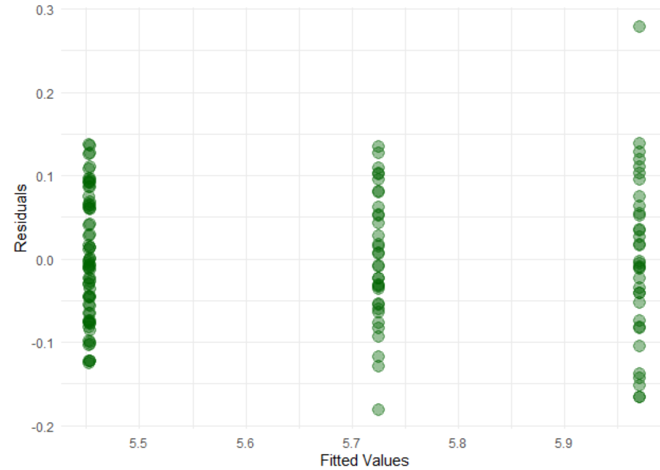


Fig. 5. Plot of residuals vs. fitted values.

TABLE II
ANALYSIS OF VARIANCE TABLE OF MAE.

Source	DF	SS	MS	F-value	**P-value**
Technique	3	6.5268	2.17562	322.31	$< 2.2e-16$
Residuals	136	0.9180	0.00675		

Three groups of statistically different means –a, b and c– were obtained. The averages of the DT and RF MAEs belong to the same group and present the smallest values; therefore, it is confirmed that they are the techniques with the lowest prediction error but there are no significant differences between them. However, the average training time of DT is 0.0651 (0.00615) seconds while that of RF is 0.445 (0.0111) seconds. Therefore, DT is as suitable as RF for the identification of the ideal animal weight but is much faster (almost 7 times).

B. Result analysis

After having found the best ideal weight-prediction technique for cattle, experimentation was carried out to evaluate the performance of the identification model in the detection of abnormal weights. The used data are of female and male crossbred cattle of the genetic groups Angus x Zebu (AC); Bon x Zebu (BC); Zebu x Angus x Zebu (CAC); Zebu x Zebu (CC); Holstein x Zebu (HC); Bon x Angus x Zebu (BAC);Romo x Angus x Zebu (RAC), typical in the Colombian farms.

Fig 6 and 7 compare the ideal weight growth curve described by the identification model of female cattle of the AC and CAC breeds with the simulated weights (point cloud).

TABLE III
MULTIPLE COMPARISON RESULTS.

Technique	DT	RF	GB	KN
Mean	5.4525	5.4536	5.7249	5.9714
Group	a	a	b	c

Since the curve represents the ideal temporal behavior of the weight, points that are far away from it are potentially anomalous, but it is necessary to define the magnitude of the distance that separates an anomalous weight from one considered normal. For this, multiple predictions of ideal weight are obtained for each combination of day, breed and gender. Ideal weight intervals (non-outliers) are then constructed to detect anomalous weights using a forest insolation algorithm, which takes less than a minute to perform this process. This allows the rapid detection of anomalous increases and decreases in animal weight at any desired time for decision making.

Since farmers frequently monitor animal weights, it is possible that on some days the weight of a particular animal will be identified by the model as an outlier and on others as normal. This is particularly important because the farmer could try to identify the reason for these abnormal weights by analyzing potential causes such as animal health or pasture quality, among others, and make decisions to achieve the animal's ideal weight.

Table IV shows some examples of animals to which the variables of interest are measured and studied if their weight is anomalous or not. In this table, the measurements of the characteristics of age, animal breed, gender and weight are presented; examples of ideal weights estimated by the identification model are shown; and the weight of each animal is classified as normal or abnormal.

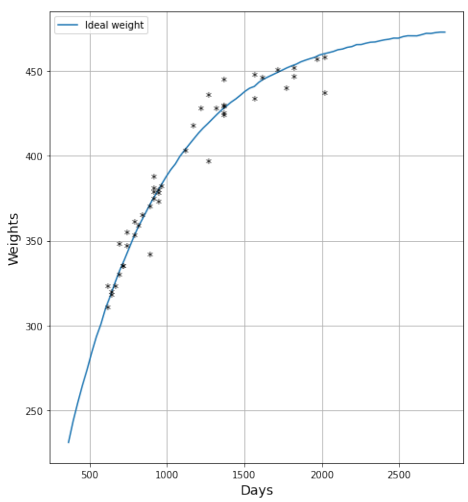


Fig. 6. Ideal weight-curve versus real weight in AC breed, female

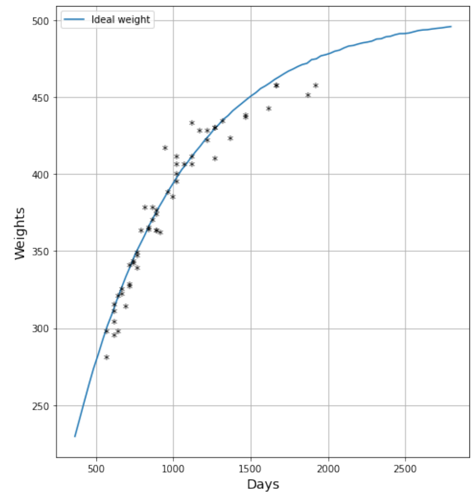


Fig. 7. Ideal weight-curve versus real weight in CAC breed, female

TABLE IV
EXAMPLE OF ANOMALOUS DATA DETECTION.

Days	Actual Weight	Breed	Gender	Example ideal Weights	State
600	361	AC	F	[332, 368, ..., 373, 387, ...]	normal
600	259	BAC	F	[309, 374, ..., 344. 368, ...]	abnormal
600	285	AC	M	[311, 363, ..., 387, 373, ...]	abnormal
678	421	BC	F	[325, 412, ..., 402, 382, ...]	abnormal

VI. Conclusion and Future Work

In this paper, we presented four regression techniques used to build an identification model capable to detect anomalies in a cattle-fattening process. The current study is the first to propose a supervised machine-learning method, using a regression technique, to build an identification model of the cattle fattening process, which can be used to detect anomalies in weight data in cattle by rotational grazing. The model uses data collected at a beef farm. The current study provides an effective solution to detect anomalous body-weight observations of cattle, using an identification model.

Future work is aimed to (i) define other identification models based on other principles (e.g., mathematical models describing the fattening process, or fattening description models based on fuzzy theory [37]), (ii) devise data-driven diagnostic models in a beef-production process, (iii) incorporate prescription models to define how to act in the event of an anomaly diagnosed in the beef-production process, (iv) Incorporate all these models in an autonomous cycle of data analysis tasks, in order to automate the supervision of the beef-production process.

Acknowledgment

All the authors would like to thank the "Vicerrectoría de Descubrimiento y Creación", of Universidad EAFIT, for their support on this research.

References

[1] D. Robinson, "Assessing the accuracy of modelling weight gain of cattle using feed efficiency data," Livestock Production Science, vol. 95, no. 3, pp. 187–200, 2005.

[2] R. García, J. Aguilar, M. Toro, A. Pinto, and P. Rodríguez, "A systematic literature review on the use of machine learning in precision livestock farming," Computers and Electronics in Agriculture, vol. 179, p. 105826, 2020.

[3] J. Aguilar and M. Cerrada, "Fuzzy classifier system and genetic programming on system identification problems," in Proceedings of the 3rd Annual Conference on Genetic and Evolutionary Computation, GECCO'01, (San Francisco, CA, USA), p. 1245–1251, Morgan Kaufmann Publishers Inc., 2001.
[4] L. Ljung, "Perspectives on system identification," Annual Reviews in Control, vol. 34, no. 1, pp. 1–12, 2010.
[5] G. Morota, R. V. Ventura, F. F. Silva, M. Koyama, and S. C. Fernando, "Big data analytics and precision animal agriculture symposium: Machine learning and data mining advance predictive big data analysis in precision animal agriculture1," Journal of Animal Science, vol. 96, pp. 1540–1550, 03 2018.
[6] S. Ma, Q. Yao, T. Masuda, S. Higaki, K. Yoshioka, S. Arai, S. Takamatsu, and T. Itoh, "Development of an anomaly detection system for cattle using infrared image and machine learning," Sens. Mater., vol. 32, no. 12, pp. 4139–4149, 2020.
[7] J. Haladjian, Z. Hodaie, S. Nüske, and B. Brügge, "Gait anomaly detection in dairy cattle," in Proceedings of the Fourth International Conference on Animal-Computer Interaction, ACI2017, (New York, NY, USA), Association for Computing Machinery, 2017.
[8] N. Wagner, V. Antoine, J. Koko, M.-M. Mialon, R. Lardy, and I. Veissier, "Comparison of machine learning methods to detect anomalies in the activity of dairy cows," in Foundations of Intelligent Systems (D. Helic, G. Leitner, M. Stettinger, A. Felfernig, and Z. W. Raś, eds.), (Cham), pp. 342–351, Springer International Publishing, 2020.
[9] Y. Chung, J. Lee, S. Oh, D. Park, H. Chang, and S. Kim, "Automatic detection of cow's oestrus in audio surveillance system," Asian-Australasian journal of animal sciences, vol. 26, no. 7, p. 1030–1037, 2013.
[10] E. Kramer, D. Cavero, E. Stamer, and J. Krieter, "Mastitis and lameness detection in dairy cows by application of fuzzy logic," Livestock Science, vol. 125, no. 1, pp. 92–96, 2009. cited By 49.
[11] H. Park, D.-H. Park, and S.-H. Kim, "Deep learning-based method for detecting anomalies of operating equipment dynamically in livestock farms," in 2020 International Conference on Information and Communication Technology Convergence (ICTC), pp. 1182–1185, 2020.
[12] Y. Cai, L. Ma, and G. Liu, "A night-time anomaly detection system of hog activities based on passive infrared detector," Applied Engineering in Agriculture, vol. 35, no. 4, pp. 481–493, 2019.
[13] A. Calera, F. Lozano, I. Campos, and A. Osann, "Selected farming services with satellite assistance: Mapping crops, optimizing water resources and assessing crop and pasture drought stress," pp. 62–72, 2013. cited By 0.
[14] S. Neethirajan and B. Kemp, "Digital livestock farming," Sensing and Bio-Sensing Research, vol. 32, 2021. cited By 0.
[15] J. Aguilar, M. Cerrada, G. Mousalli, F. Rivas, and F. Hidrobo, "A multiagent model for intelligent distributed control systems," Lecture Notes in Computer Science, vol. 3681, pp. 191–197, 2005.
[16] J. Aguilar, M. Cerrada, and F. Hidrobo, "A methodology to specify multiagent systems," in Agent and Multi-Agent Systems: Technologies and Applications (N. T. Nguyen, A. Grzech, R. J. Howlett, and L. C. Jain, eds.), (Berlin, Heidelberg), pp. 92–101, Springer Berlin Heidelberg, 2007.
[17] J. Vizcarrondo, J. Aguilar, E. Exposito, and A. Subias, "Mape-k as a service-oriented architecture," IEEE Latin America Transactions, vol. 15, no. 6, pp. 1163–1175, 2017.
[18] H. Iba, T. Sato, and H. de Garis, "System identification approach to genetic programming," in Proceedings of the First IEEE Conference on Evolutionary Computation. IEEE World Congress on Computational Intelligence, pp. 401–406 vol.1, 1994.
[19] J. Aguilar, M. Cerrada, E. Colina, and A. Titli, "Dynamical adaptive fuzzy systems: An application on system identification," IFAC Proceedings Volumes, vol. 35, no. 1, pp. 403–408, 2002. 15th IFAC World Congress.
[20] J. Waissman, R. Sarrate, T. Escobet, J. Aguilar, and B. Dahhou, "Wastewater treatment process supervision by means of a fuzzy automaton model," in Proceedings of the 2000 IEEE International Symposium on Intelligent Control. Held jointly with the 8th IEEE Mediterranean Conference on Control and Automation (Cat. No.00CH37147), pp. 163–168, 2000.
[21] C. A. Carabalí., L. Tituaña., J. Aguilar., O. Camacho., and D. Chavez., "Inverse response systems identification using genetic programming," in Proceedings of the 14th International Conference on Informatics in Control, Automation and Robotics - Volume 1: ICINCO,, pp. 238–245, INSTICC, SciTePress, 2017.
[22] J. Aguilar, "The evolutionary programming in the identification of discreet events dynamic systems," IEEE Latin America Transactions, vol. 5, no. 5, pp. 301–310, 2007.
[23] M. Araujo, J. Aguilar, and H. Aponte, "Fault detection system in gas lift well based on artificial immune system," in Proceedings of the International Joint Conference on Neural Networks, 2003., vol. 3, pp. 1673–1677 vol.3, 2003.
[24] T. Weber, J. Sossenheimer, S. Schäfer, M. Ott, J. Walther, and E. Abele, "Machine learning based system identification tool for data-based energy and resource modeling and simulation," Procedia CIRP, vol. 80, pp. 683–688, 2019. 26th CIRP Conference on Life Cycle Engineering (LCE) Purdue University, West Lafayette, IN, USA May 7-9, 2019.
[25] S. García, J. Luengo, and F. Herrera, "Tutorial on practical tips of the most influential data preprocessing algorithms in data mining," Knowledge-Based Systems, vol. 98, pp. 1–29, 2016.
[26] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," Journal of Machine Learning Research, vol. 13, no. 10, pp. 281–305, 2012.
[27] Y. Zhao, Z. Nasrullah, and Z. Li, "Pyod: A python toolbox for scalable outlier detection," arXiv preprint arXiv:1901.01588, 2019.
[28] L. C. Brito, G. A. Susto, J. N. Brito, and M. A. Duarte, "An explainable artificial intelligence approach for unsupervised fault detection and diagnosis in rotating machinery," Mechanical Systems and Signal Processing, vol. 163, p. 108105, 2022.
[29] A. K. Rai and R. K. Dwivedi, "Fraud detection in credit card data using unsupervised machine learning based scheme," in 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 421–426, IEEE, 2020.
[30] L. Meneghetti, M. Terzi, S. Del Favero, G. A. Susto, and C. Cobelli, "Data-driven anomaly recognition for unsupervised model-free fault detection in artificial pancreas," IEEE Transactions on Control Systems Technology, vol. 28, no. 1, pp. 33–47, 2018.
[31] T. Barbariol, E. Feltresi, and G. A. Susto, "Self-diagnosis of multiphase flow meters through machine learning-based anomaly detection," Energies, vol. 13, no. 12, p. 3136, 2020.
[32] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in 2008 eighth ieee international conference on data mining, pp. 413–422, IEEE, 2008.
[33] L. Rokach and O. Maimon, Decision Trees, pp. 165–192. Boston, MA: Springer US, 2005.
[34] G. Biau and B. Cadre, Optimization by Gradient Boosting, pp. 23–44. Cham: Springer International Publishing, 2021.
[35] J. Korstanje, The kNN Model, pp. 169–177. Berkeley, CA: Apress, 2021.
[36] R. Genuer and J.-M. Poggi, Random Forests, pp. 33–55. 09 2020.
[37] E. Puerto, J. Aguilar, C. López, and D. Chávez, "Using multilayer fuzzy cognitive maps to diagnose autism spectrum disorder," Applied Soft Computing, vol. 75, pp. 58–71, 2019.

Appendix D

Supervision system of the fattening process of cattle in rotational grazing using fuzzy classification

Supervision System of the Fattening Process of Cattle in Rotational Grazing using Fuzzy Classification

Charles Benitez
Faculty of Engineering, Universidad de los Andes
Universidad de los Andes
Merida, Venezuela
0000-0001-9975-4790

Rodrigo García
GIDITIC, Universidad EAFIT
Faculty of Engineering, Universidad del Sinú
Medellin, Colombia
0000-0002-5935-8862

Jose Aguilar
GIDITIC, Universidad EAFIT
Universidad de Alcalá, Dept. Automática
Colombia, Venezuela, España
0000-0003-4194-6882

Marvin Jiménez
Faculty of Engineering
Universidad del Sinú
Montería, Colombia
0000-0002-0026-3412

Horderlin Robles
Faculty of Engineering
Universidad del Sinú
Montería, Colombia
0000-0002-7705-7334

2022 XVLIII Latin American Computer Conference (CLEI) | 978-1-6654-7671-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/CLEI56649.2022.9959950

***Resumen*—Cattle breeding has been one of the most important industrial sectors in the world, since it is related to food security and the survival of the human race. Cattle diagnostics is a fundamental procedure for cattle breeders because it allows them to make strategic decisions, such as timely treatment in case of any abnormality (e.g., weight gain in herds, in their paddocks). This article aims to present a system to diagnose weight loss or gain in cattle under a rotational grazing scheme, considering the health status of the animal and the pasture. The diagnostic system is based on a fuzzy classifier that uses fuzzy logic to define the rules that characterize the diagnostic process, and fuzzy reasoning to determine the current situation given an input. In addition, the fuzzy classifier optimizes the rules using genetic algorithms, which modify the membership functions, providing a more accurate system for diagnosis. We tested our proposal with experimental cases, with promising results. The accuracy metrics have high values, indicating a low error rate in terms of false positives. In general, the values of the quality metrics are very good, with an accuracy close to 100 % and an *Area Under the Curve* close to 1.**

***Index Terms*—Diagnostic System, Precision Livestock Farming, Rotational Grazing, Fuzzy Classifier, Genetic Algorithms**

I. INTRODUCCIÓN

La actividad ganadera en las naciones ubicadas entre trópicos presenta cada vez una mayor exigencia productiva debido al crecimiento demográfico, por lo que esta actividad debe constantemente asumir dinámicas de transformación y adecuación. Por otro lado, las estaciones climáticas intertropicales se caracterizan por contar con estaciones de invierno y verano, los cuales condicionan la actividad ganadera en cuanto a la alimentación de los animales, por lo que se requiere mejorar los procesos alimenticios considerando estos aspectos, para que esta actividad genere mayores volúmenes de ganancia. A su vez, la industria ganadera presenta una serie de dificultades para detectar fallos en sus sistemas de producción, lo cual afecta directamente el crecimiento ideal del ganado y la producción óptima de los potreros. La problemática radica en la incapacidad de poder tomar decisiones en tiempo real, usando los datos, de lo que puede estar ocurriendo en el proceso de engorde del ganado. Por lo anterior, se requiere desarrollar un sistema de diagnóstico que actúe bajo incertidumbre, con el objetivo de ayudar en las decisiones de los ganaderos para aumentar la eficiencia productiva.

Algunos trabajos relacionados con nuestra propuesta son los siguientes. Palomino y Loza [1] diseñaron un sistema de pastoreo rotacional para un hato lechero alto andino. La finalidad de ese proyecto era mejorar los sistemas de pastoreo tradicionales para ganado lechero alto andino de doble propósito. Para el diseño se usó el software SAS Planet, y se plotearon planos del predio y de las áreas con fines de agricultura y vivienda, con el fin de determinar el número ideal de potreros y la carga animal óptima de los mismos. Varios autores [2], [3] reseñan los aspectos esenciales para desarrollar una ganadería baja en emisiones de carbono. Consideraron un conjunto de efectos relacionados con el cambio climático, y su repercusión sobre la relación suelo-planta-animal. Proponen un sistema de pastoreo intensivo que garantiza una ganadería baja en emisiones, lo cual representa una opción ante el cambio climático. A su vez, García y otros [4] proponen un enfoque para detectar anomalías en el proceso de engorde del ganado. Este enfoque usa el registro histórico real del peso de los animales, para identificar si los animales han ganado el peso adecuado a lo largo del tiempo. Ellos comparan varias técnicas de aprendizaje automático (Árbol de Decisión, *Gradient Boosting*, regresión basada en *K-Nearest Neighbors y Random Forest*) en la tarea de detección de pesos anómalos, usando como métricas de calidad el Error Absoluto Medio.

Por otro lado, en cuanto a sistemas clasificadores difusos, Ramírez y otros [5] diseñaron un sistema clasificador difuso para el establecimiento de los estados funcionales de un sistema de producción de aire medicinal. Para el diseño y sintonización del clasificador difuso se utilizó el histórico de datos del proceso, para identificar todos los estados funcionales útiles para el monitoreo del proceso. El establecimiento de estados funcionales a partir de la clasificación difusa permite la programación de acciones correctivas en el proceso a partir del diagnóstico realizado. Finalmente, en [6] aplicaron la lógica difusa y los algoritmos genéticos para la determinación de tratamientos en enfermedades neoplásicas malignas. El artículo propone un clasificador difuso optimizado utilizando un algoritmo genético híbrido con una técnica de agrupamiento difuso. Ellos implementan un prototipo y lo evalúan en datos sintéticos de tratamiento contra enfermedades neoplásicas malignas.

Este trabajo tiene como objetivo principal desarrollar un sistema para diagnosticar el progreso individual de engorde, y las causas de pérdida o ganancia de peso, del ganado en pastoreo rotacional. El sistema esta basado en un clasificador difuso, que analiza los datos y arroja un diagnóstico. El sistema usa reglas para analizar la situación actual del peso del ganado. Además, el sistema optimiza las reglas para adecuarlas a los datos reales del ganado bajo supervisión. La organización de este trabajo es la siguiente. La sección 2 presenta el marco teórico usado en este trabajo. La sección 3 muestra el diseño de nuestro sistema de diagnóstico. Después, la sección 4 describe los experimentos realizados con el sistema, para evaluar su calidad. A continuación, la Sección 5 compara este trabajo con otros anteriores. Finalmente, la Sección 5 presenta las conclusiones.

II. Marco teórico

II-A. *Sistema Clasificador Difuso*

En general, los sistemas difusos ha demostrado ser de gran utilidad para representar el comportamiento o dinámica de los sistemas mediante reglas difusas. Tradicionalmente, estos sistemas se basan en la información suministrada por expertos; sin embargo, en sistemas complejos, las reglas así construidas no permitían una simulación exacta del sistema [7]. La búsqueda de sistemas difusos que se adapten a la dinámica de sistemas complejos, ha conllevado al desarrollo de investigaciones en técnicas de extracción de reglas difusas a partir de datos de entrada y salida [8]–[10]. Los algoritmos de clasificación difusa representa una de las técnicas para el desarrollo de sistemas difusos adaptativos [11], [12].

Un Sistema Clasificador difuso está compuesto por dos grandes componentes. Un *subsistema de reglas* que permite clasificar la información de entrada. Ese subsistema define las reglas del tipo *Si-Entonces* usando variables difusas, y utiliza un razonador difuso para inferir una conclusión. El segundo componente de un sistema clasificador difuso es el *subsistema adaptativo*. Tradicionalmente, ese subsistema se basa en algoritmos genéticos (AG), que se inspiran en la evolución biológica como estrategia para resolver problemas de optimización [13]. En particular, el clasificador difuso lo usa para optimizar las reglas, y de esta manera, su funcionamiento.

II-B. *Ganadería de precisión*

La ganadería de precisión es uno de los sectores que más innovaciones ha incorporado en los últimos años, gracias al desarrollo de tecnologías orientadas a la mejora de la eficiencia en las explotaciones ganaderas y de la calidad de los productos de origen animal. Estas tecnologías permiten optimizar los recursos, aumentar el rendimiento, controlar el impacto ambiental, y mejorar el bienestar de los animales. Para ello, se requiere el monitoreo de la salud animal y producción, la medición de indicadores biométricos de carácter fisiológico y morfológico, entre otras cosas [14].

En el caso de procesos de diagnóstico, el uso de ganadería de precisión permite el desarrollo de una tecnología de supervisión amigable, no intrusiva, permitiendo la interacción entre expertos y productores. Particularmente, los sistemas de diagnóstico en la ganadería de precisión buscan disminuir las pérdidas y mejorar la producción a largo plazo, a través de dinámicas de seguimiento y monitoreo al ganado. Por ejemplo, un sistema de diagnóstico podría evaluar información cuantitativa para caracterizar a los animales de forma rápida y adecuada, mediante el estudio individual y/o por lotes de animales. El sistema realizaría la interpretación de la salud animal y determinaría el rendimiento del potrero, para identificar los patrones en el proceso productivo, y a partir de allí, poder diagnosticar las posibles causas de un bajo o alto engorde del ganado [4].

III. Diseño del Sistema de Diagnóstico

En esta sección se hablará del diseño y funcionamiento del sistema de diagnóstico basado en un clasificador difuso.

III-A. *Especificación del Sistema Clasificador Difuso*

La arquitectura del clasificador difuso se puede apreciar en la Figura 1. En este esquema se observan 2 bloques, el primer bloque, llamado *sistema de diseño y optimización de reglas difusas*, y el segundo lleva por nombre *sistema difuso*. En las Figuras 2 y 3, respectivamente, se detallan ambos sistemas.

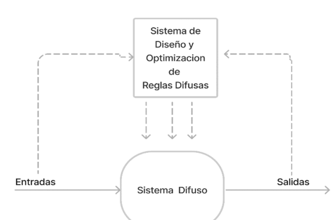


Figura 1: Arquitectura del clasificador difuso.

La Figura 2 muestra el sistema de diseño y optimización de las reglas difusas. Las reglas son diseñadas y creadas mediante un algoritmo de agrupamiento difuso. El presente trabajo utiliza el algoritmo *fuzzy-c-means* (FCM), el cual permite agrupar datos difusos (cada grupo determina un conjunto difuso para la variable lingüística/difusa de interés) [1], [15]–[18]. Por otro lado, para la optimización se usa un AG, cuya tarea es mejorar la definición de los conjuntos difusos (adapta la función de membresía de cada conjunto difuso).

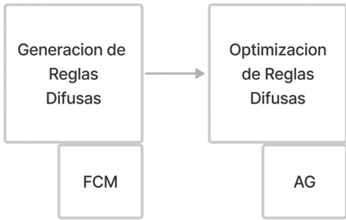


Figura 2: Sistema de Diseño y Optimización de reglas difusas

En la Figura 3 se muestra el sistema difuso. Sus componentes son el fusificador de las entradas, el motor de inferencia (razonador difuso), el generador de salida (defusificador), y la base de conocimiento (las reglas difusas), la cual es actualizada cada cierto tiempo por el sistema de optimización de reglas difusas mostrado en la Figura 2.

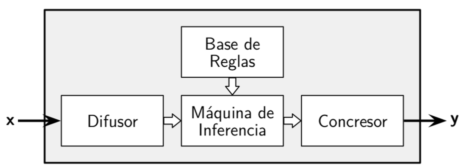


Figura 3: Sistema difuso

III-B. *Especificación del sistema difuso*

III-B1. Sistema Difuso: El corazón de este sistema se centra en la definición de las variables lingüísticas que conforman el modelo de diagnóstico. Este modelo está constituido por un conjunto de reglas difusas que usan las variables lingüísticas. Así, el sistema difuso representa el modelo de diagnóstico propuesto en este trabajo.

La primera etapa para diseñar el sistema difuso consiste en crear las etiquetas (conjuntos difusos) necesarias para cada variable lingüística. El algoritmo FCM permite definir los conjuntos difusos a partir de la matriz de pertenencia que él genera para cada variable lingüística [1], [19]. Para ello, FCM determina los grupos difusos presentes en una variable lingüística (serán los conjuntos difusos), con el grado de pertenencia de los valores de la variable en cada uno de ellos. Los grados de pertenencia están en un rango de $[0, 100]$, y representan la pertenencia parcial a cada conjunto difuso (clase) de cada valor de cada variable lingüística. A continuación, se describen las variables lingüísticas, y sus conjuntos difusos, de nuestro sistema de diagnóstico difuso.

Las *variables lingüísticas* utilizadas son las vinculadas al proceso de engorde. Las variables de entrada son la edad, el Potrero y el diferencial del peso final con el peso inicial; y la variable de salida es el estado del animal.

Edad: su universo de discurso es $[0, 30]$, y representa los meses de vida. Sus conjuntos difusos son [ternero, novillo, adulto]. En la Figura 4 se muestra cada conjunto difuso.

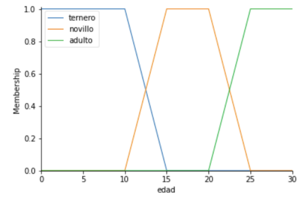


Figura 4: Variable Edad

Diferencial: su universo de discurso es $[0, 100]$, y representa un porcentaje de diferencia entre el peso inicial y el actual. Sus conjuntos difusos son $[bajo, medio, alto]$. En la Figura 5 se muestra cada conjunto difuso.

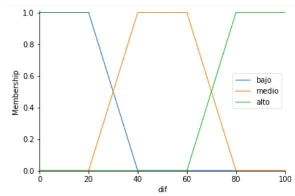


Figura 5: Variable Diferencial

Potrero: su universo de discurso es $[0, 10]$, y representa la puntuación de rendimiento del potrero, caracterizada por la calidad del forraje y la tolerancia del forraje a la situación climática. Sus conjuntos difusos son $[malo, regular, bueno]$. En la Figura 6 se muestra cada conjunto difuso.

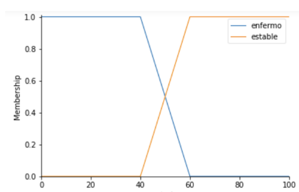


Figura 6: Variable Potrero

Estado del Animal: su universo de discurso es $[0, 100]$, representado por un porcentaje para indicar que tan bien está el animal. Sus conjuntos difusos son $[estable, enfermo]$. En la Figura 7 se muestra cada conjunto difuso.

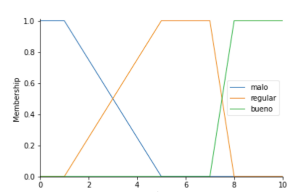


Figura 7: Variable Estado Animal

III-B2. Definición de las reglas de diagnóstico: Las reglas que modelan el sistema son de la forma $IF < \ldots > THEN < \ldots >$, En los siguientes cuadros definimos el conjunto de reglas.

Cuadro I: Base de Reglas con 2 antecedentes

		Diferencial		
		Bajo	Medio	Alto
Edad	Ternero	Enfermo	Enfermo/Estable	Estable
	Novillo	Enfermo	Enfermo/Estable	Estable
	Adulto	Enfermo/Estable	Enfermo/Estable	Estable

Cuadro II: Base de Reglas con 3 antecedentes

		Diferencial				
		Bajo	Medio	Alto		
Edad	Ternero	Enfermo	Enfermo/Estable	Estable	Regular/Malo	Potrero
	Novillo	Enfermo	Enfermo/Estable	Estable		
	Adulto	Enfermo/Estable	Enfermo/Estable	Estable		

Se tienen dos tipos diferentes de reglas difusas, dependiendo si se tiene información sobre el rendimiento del potrero. Así, se tiene un grupo de reglas con dos antecedentes (ver Cuadro I) o con tres antecedentes (ver Cuadro II). Particularmente, en el Cuadro I los dos antecedentes son las variables diferencial y edad; y en el Cuadro II ellas mismas, pero ahora también la variable potrero (los valores de ella que influyen directamente en el engorde animal, tales como regular/malo). Un ejemplo de regla para el ultimo caso es:

Si la edad es ternera y el diferencial es bajo y el potrero es malo, entonces el estado del animal es enfermo

Estas reglas serán ajustadas/optimizadas al conjunto de datos de entrada usando AGs. Específicamente, las funciones de membresías de los conjuntos difusos serán las ajustadas según los datos de entrada.

III-C. Características del Algoritmo Genético

III-C1. Especificación y evolución de los individuos: Estructura de los individuos: Un factor de especial interés en el diseño del proceso evolutivo para optimizar el clasificador difuso es el esquema de representación usado para codificar cada una de las posibles soluciones. Para esta investigación, el individuo es definido de la siguiente manera:

Cromosoma = límites de la función de pertenencia de los conjuntos difusos de cada variable lingüística.

En este enfoque, cada individuo representa por sí mismo una solución completa, al codificar los conjuntos difusos de las variables de entrada y salida. Específicamente, cada gen representa un vértice de la función de pertenencia de los conjuntos de cada variable lingüística, suponiendo una forma trapezoidal para la definición de ellas.

La representación de cada variable lingüística sería de la siguiente manera (vértices de las funciones de pertenencia de sus conjuntos difusos):

$$Edad = [10, 15, 20, 25] = [e1, e2, e3, e4]$$
$$Diferencial = [20, 40, 60, 80] = [d1, d2, d3, d4]$$
$$Potrero = [1, 5, 7, 8] = [p1, p2, p3, p4]$$
$$Estado = [40, 60] = [c1, c2]$$

Así, la codificación del cromosoma sería la concatenación de la descripción de la función de pertenencia de los conjuntos difusos de cada variable. Por otro lado, dado que hay reglas que no contienen la variable difuso potrero, existen 2 casos para la codificación de los cromosomas, el primero no contiene los conjuntos difusos de la variable difusa potrero, y el segundo caso si los contiene. Así, en el primer caso, considerando el orden de los genes, tal que primero se colocan los antecedentes (edad y diferencial) y después el consecuente (estado), los individuos quedan de la siguiente forma (ver Figura 8):

e1	e2	e3	e4	d1	d2	d3	d4	c1	c2

Figura 8: Cromosoma para el AG en el caso 1

En el segundo caso de reglas difusas, en el que se considera al rendimiento del potrero en el análisis, en los antecedentes quedarían edad, diferencial y potrero, y en el consecuente estado (ver Figura 9).

e1	e2	e3	e4	d1	d2	d3	d4	p1	p2	p3	p4	c1	c2

Figura 9: Cromosoma para el AG en el caso 2

III-C2. Evolución de los individuos: Para la optimización de las reglas difusas con AG, se debe generar de forma aleatoria una población inicial de individuos. El objetivo del AG es de modificar los conjuntos difusos (sus funciones de pertenencia), de tal manera de adecuar las reglas a los datos. El proceso evolutivo es el siguiente.

- Se seleccionan individuos para la reproducción mediante el siguiente esquema: se ordenan los individuos de mayor a menor aptitud, y se descartan los de aptitud menor a 0.7. La función de aptitud es definida en la siguiente sección.
- Posteriormente, se usan operadores genéticos para generar nuevos individuos. En particular, se usan los operadores de cruzamiento y mutación, los cuales se implementaron de la siguiente manera:
  1. Cruzamiento: se eligen dos cromosomas y el punto de corte aleatoriamente, para realizar el cruzamiento entre los dos.
  2. Mutación: se escoge un cromosoma, y se modifica uno de sus genes, escogido aleatoriamente.

- En cada iteración se genera un número de descendientes igual al tamaño inicial de la población, y se reemplazan los peores individuos en la población por los mejores nuevos, según la función de aptitud.

Al término de esto, si se llega al óptimo o se cumple la condición de parada, el individuo con mejor valor en la función de aptitud es seleccionado para actualizar los vértices de los conjuntos difusos de las variables difusas.

III-D. Función de Aptitud

Se utilizó una función de aptitud basada en el cálculo de las siguientes medidas:

Score: Mide la precisión de un individuo. En este caso, como es basado en datos etiquetados sobre diagnóstico, el score establece si el consecuente es cierto (regla se activa) cuando debería. Se calcula mediante la siguiente ecuación:

$$Score = \frac{pv}{(pv + pf)} \tag{1}$$

Donde pv es el número de verdaderos positivos, y pf el número de falsos positivos. Particularmente, se calcula el score para cada una las reglas para el conjunto de datos de entrada, y se promedia.

Grado de certeza o de disparo de cada regla: el cual viene siendo el nivel de activación de una regla, calculado como el producto de las funciones de membresía de los antecedentes. Si el valor es alto quiere decir que el nivel de activación de la regla es alto.

Finalmente, la función de aptitud de cada individuo es el promedio del score de cada una las reglas multiplicado por su grado de certeza. Así, el proceso evolutivo, usando esta función de aptitud híbrida (score y grado de certeza), permite ajustar los conjuntos difusos para que se adapten a los datos del contexto.

IV. EXPERIMENTACIÓN

En esta sección se analiza el comportamiento de nuestro sistema. Para ello, se desarrolla un protocolo experimental que describe el contexto donde se prueba el sistema clasificador difuso, y se evalúa su comportamiento mediante métricas de calidad.

IV-A. Contexto Experimental

IV-A1. Simulador: Para los experimentos se utiliza un simulador ganadero (ver https://github.com/devraxielh/Simulador_Ganadero), el cual emula todas las condiciones climáticas y del suelo de una finca, como también, del comportamiento de engorde del ganado. Particularmente, el simulador genera los valores de las variables que describen el terreno al cual se le aplica pastoreo rotacional, y de la población de ganado que se está engordando. De esta manera, el simulador genera los datos aleatoriamente que alimentan el sistema clasificador difuso. Los datos de interés dados por el simulador son los siguientes: número de animales, edad de cada animal, peso inicial de cada animal, evolución del peso de cada animal, número de potreros, forraje de cada potrero, estación climática de cada potrero, y tolerancia del forraje a climas calientes.

IV-A2. Métricas de Calidad: Las métricas para valorar la calidad del modelo de diagnóstico tienen como objetivo estimar la precisión del modelo en datasets de pruebas (diferentes a los usados en el proceso adaptativo de las reglas). Estas son:

- Precisión (score): Es el porcentaje de acierto correctos dados por el sistema adaptativo.
- Grado de certeza: Es el valor que se obtiene de la multiplicación de las funciones de membresía de los antecedentes, para determinar el grado de disparo de cada regla y deducir qué regla se activa más que otra.
- Curva ROC (Característica Operativa del Receptor): Esta es una gráfica que muestra la sensibilidad y especificidad de un modelo:
  - Sensibilidad: la probabilidad de que el modelo prediga un resultado positivo para una observación cuando en realidad el resultado es positivo.
  - Especificidad: La probabilidad de que el modelo prediga un resultado negativo para una observación cuando en realidad el resultado es negativo.
- Área Bajo la Curva (AUC): Es un valor numérico dado por el área de la curva que genera la gráfica de ROC. Cuanto mayor sea el área cubierta, mejores serán los modelos de aprendizaje automático para distinguir las clases dadas. El valor ideal para AUC es 1.

IV-B. Escenarios de Prueba

En esta sección se presentan los casos de estudio que serán analizados por el modelo de diagnóstico. Cada caso de estudio representa situaciones donde se hace prevalecer situaciones climáticas que afecten o no al potrero, también se consideran situaciones de salud animal. A continuación, se describen los casos de estudio considerados.

IV-B1. Caso de estudio 1:

IV-B1a. Objetivo: El primer caso de estudio es el óptimo, donde todos los animales se encuentran estables y el potrero es el idóneo para su uso. Las reglas que serán activadas son las del grupo de 2 antecedentes, ya que el caso es ideal, entonces el rendimiento del potrero será bueno y no afecta de manera negativa el engorde animal.

IV-B1b. Entrada al sistema: En el cuadro III presenta una vista parcial de los datos históricos de engorde del ganado (las primeras 4 columnas). Él muestra las variables de entrada (edad, peso inicial y peso final), pero también cuenta con un identificador para cada animal. Por otro lado, el peso inicial indica el peso con el que entran al potrero y el peso final representa el peso con el que salen del potrero. Tomando en cuenta que el engorde diario animal varía en un promedio de 300 a 500 gramos por día, se puede decir que el engorde es normal si al cabo de 30 días (tiempo de estadía en cada potrero) aumentan de 8 a 15 kilos aproximadamente. Esto determina que el diferencial de los pesos está dentro del rango, y por ello es un caso ideal.

Cuadro III: Representación de datos del ganado y clasificación del estado Animal en el experimento 1

Edad (meses)	Peso inicial(kg)	Peso fin(kg)	ID animal	Predicción	Grados pertenencia
19	373.0205	384.4683	V1	1	0.9171
29	389.0205	400.4683	V2	1	0.9171
23	368.0205	379.4683	V3	1	0.9171
18	402.0205	413.4683	V4	1	0.9171
22	385.0205	396.4683	V5	1	0.9171
19	404.0205	415.4683	V6	1	0.9171
22	369.0205	380.4683	V7	1	0.9171
22	407.0205	418.4683	V8	1	0.9171
19	400.0205	411.4683	V9	1	0.9171
26	418.0203	427.1916	V10	1	0.9171

El cuadro III muestra en la columna grado de pertenecía de cada dato al grupo de la etiqueta estable, por ende, la columna de predicción presenta la etiqueta binaria 1 (como estable). Esas dos columnas representan la actual predicción que realiza el sistema de clasificación para el conjunto de datos de entrada.

Cuadro IV: Rendimiento del potrero en el experimento 1

Nombre	Clima	Potrero	Forraje	Estado
Humidicola comum Tuly o Quicuio da Amazönia	Lluvia	6	7	Bueno

El cuadro IV expone los datos de entrada del potrero. Se muestra el potrero ocupado, el nombre del forraje usado en el potrero y la situación climática.

El cuadro IV muestra la columna forraje, la cual representa la puntuación que tiene el potrero en un rango de [0,10], siendo 10 la puntuación más alta. Este dato se obtiene dependiendo de la tolerancia del forraje a la situación climática a la que se encuentra. La columna estado representa el rendimiento del potrero según la interpretación difusa para categorizar el rendimiento del potrero.

IV-B1c. Análisis de resultados: Como se pudo ver, con las reglas difusas existentes, el sistema clasificador difuso infiere que todos los animales están estables para esos datos de entradas. Una vez se han realizado varias iteraciones del sistema difuso, el sistema adaptativo de las reglas es invocado (ver las siguientes figuras que muestran la adaptación de las reglas a los datos).

Se puede observar en la Figura 10 como se ajustan los conjuntos difusos de cada variable difusa usando el dataset del cuadro III. Cada uno representa el mejor individuo para las primeras 3 iteraciones del AG. Por ejemplo, vemos en la Figura 11.1 los conjuntos difusos de los antecedentes y el consecuente del mejor individuo para la primera iteración. La

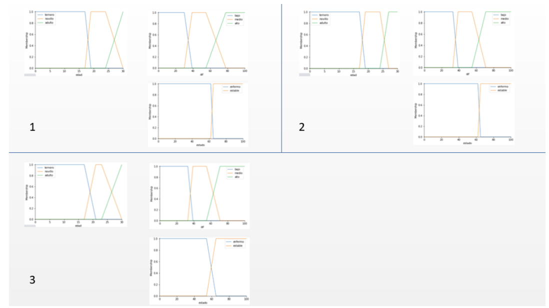


Figura 10: Adaptación de las funciones de pertenencia

variable difusa del potrero no se activa, ya que el rendimiento del potrero es bueno. Las reglas difusas que se activan con estos ajustes, usando el dataset del cuadro III, se muestran en la Figura 11.1. En la Figura 11.2 se muestra el mejor individuo de la siguiente iteración, y su base de reglas en la Figura 11.2. Finalmente, la Figura 11.3 muestra el mejor individuo cuando se detiene el AG, bien sea porque se cumple la condición de parada o porque encontró el óptimo. La base de regla final mostrada en la Figura 11.3 son las reglas necesarias del sistema para el dataset del cuadro III. El individuo de la Figura 11.3 representa la potencial solución, y al comparar las reglas activadas finales (Figura 11.3) con el conjunto de datos mostrados en el cuadro III que etiqueta a todos los animales estables, vemos que coherentemente las reglas activadas tienen en el consecuente solo la etiqueta de estable.

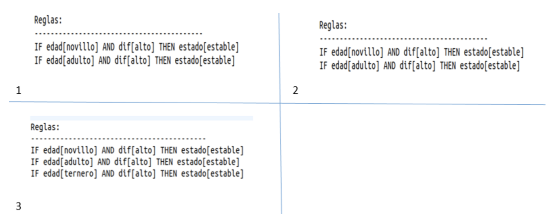


Figura 11: Evolución de las reglas activadas en el caso 1

Cuadro V: Métricas del caso 1.

Métricas	**Valores**
Precisión	100 %
Certeza de reglas	R1:0.85, R2: 0.88,R3: 0.89
AUC	1

Finalmente, con las reglas difusas finales (ver Figura 11) se procede a diagnosticar usando el dataset de prueba. Con el resultado del proceso de inferencia de cada individuo, se pasa a calcular los promedios de las métricas de calidad del sistema (ver cuadro V y Figura 12).

La curva ROC nos ayuda a identificar el rendimiento del modelo, y permite calcular el AUC, que sería el área bajo la curva. Notamos en el cuadro V que AUC vale 1 y la precisión es de 100 %. Recordemos que la certeza de cada regla (Ri) viene determinada por su grado de disparo, el cual varía entre 0 y 1. Vemos que las 3 reglas tienen un alto nivel de disparo (mayores o iguales a 0.85). Los anteriores resultados nos dicen que el modelo es robusto y realiza una perfecta clasificación.

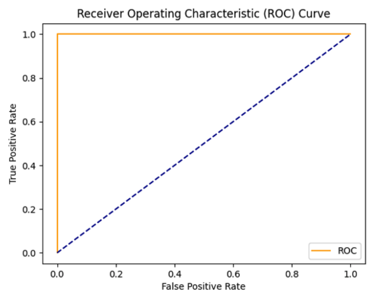


Figura 12: Curva ROC caso 1

IV-B2. Caso de estudio 2:

IV-B2a. Objetivo: El segundo caso de estudio presenta una variación, donde al menos un animal es etiquetado enfermo y el rendimiento del potrero es bueno, a pesar de la estación climática. Las reglas que serán activadas son las del grupo de 2 antecedentes, ya que el rendimiento del potrero será bueno y no afecta de manera negativa el engorde animal.

IV-B2b. Entrada al sistema: Los datos de entrada son los siguientes:

Cuadro VI: Representación de datos del ganado y clasificación del estado Animal en el experimento 2

Edad (meses)	Peso inicial(kg)	Peso fin(kg)	ID animal	Predicción	Grados pertenencia
19	332.0	334.5431	V1	0	0.2543
29	348.0	356.5431	V2	1	0.8543
23	327.0	335.5431	V3	1	0.8543
18	361.0	369.5431	V4	1	0.8543
22	344.0	352.5431	V5	1	0.8543
19	363.0	371.5431	V6	1	0.8543
22	328.0	336.5431	V7	1	0.8543
22	366.0	374.5431	V8	1	0.8543
19	359.0	367.5431	V9	1	0.8543
26	380.0	388.1348	V10	1	0.8134

El cuadro VII contiene los datos utilizados para la entrada del potrero. Se muestra el potrero ocupado, el nombre del forraje usado en el potrero, y la condición climática.

Cuadro VII: Rendimiento del potrero en el experimento 2.

Nombre	Clima	Potrero	Forraje	Estado
Humidicola comum Tuly o Quicuio da Amazônia	Seco	1	7	Bueno

Ese mismo cuadro VII muestra la columna forraje, la cual representa calidad del potrero. La columna estado presenta el rendimiento del potrero.

IV-B2c. Análisis de resultados: Con las reglas difusas existentes, vimos que el sistema clasificador difuso infiere que solo un animal está enfermo para esa entrada dada (ver tabla VI). Una vez se han realizado varias iteraciones del sistema difuso, el sistema adaptativo de las reglas es invocado.

Se observa en la Figura 14 la evolución del mejor individuo a través de las generaciones del AG. Las reglas activadas usando el mejor individuo de la 1ra generación (ver Figura 14.1) y el dataset de entrenamiento, son mostradas en la Figura 13.1, pero estas reglas no satisfacen el sistema, ya que solo tiene reglas con consecuente estable. En la 2da iteración, en la Figura 14.2, sus reglas activadas con el dataset de entrenamiento se pueden ver en la Figura 13.2. Ya en esa iteración se nota una mejora en las reglas porque aparece en el consecuente el estado enfermo.

Finalmente, en la 4ta iteración se llega a la condición de parada, y su mejor individuo representa la solución final (ver Figura 14.4). En este caso, las reglas finales (ver Figura 13.4) son las necesarias para el sistema de diagnóstico para el conjunto de datos mostrados en el cuadro VI.

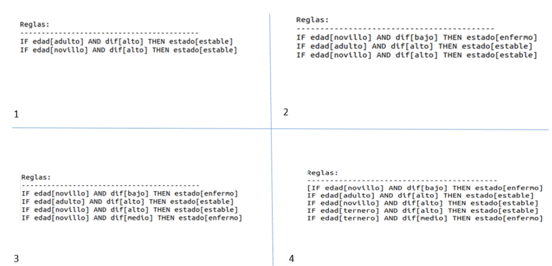


Figura 13: Evolución de las reglas que se activan en el experimento 2.

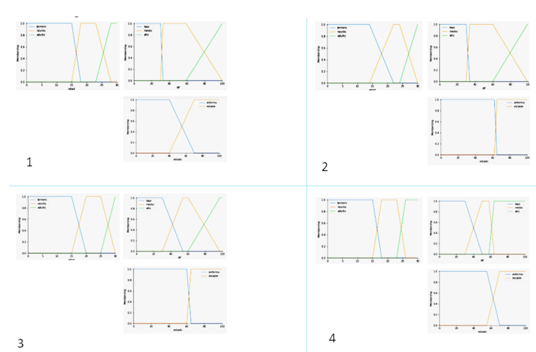


Figura 14: Adaptación de las funciones de pertenencia.

Finalmente, con las reglas difusas finales (ver Figura 13.4) se procede a probar el sistema de diagnóstico usando los datos de prueba. Con el resultado del proceso de inferencia, se calculan las métricas de calidad del sistema (ver cuadro VIII y Figura 15).

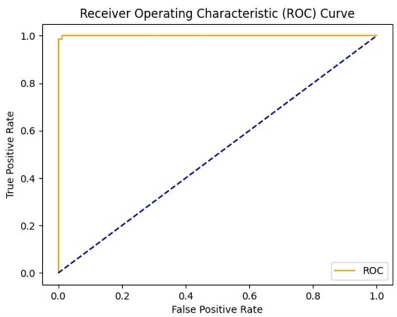


Figura 15: Curva ROC del caso 2.

Cuadro VIII: Métricas del caso 2.

Métricas	**Valores**
Precisión	97 %
Certeza de reglas	R1:0.85, R2:0.88,R3:0.90,R4:0.85,R5:0.85
AUC	1

Notamos en el cuadro VIII que AUC vale 0.98, y la precisión 97 %. Por otro lado, la certeza de cada regla difusa final es igual o mayor a 0.85. Todo lo anterior nos dice que el modelo es robusto, realiza un buen diagnóstico, y sus reglas son adecuadas para los datos de entrada.

IV-B3. Caso de estudio 3:

IV-B3a. Objetivo: El tercer caso de estudio presenta una variación, el cual consiste en analizar el engorde animal, tomando en cuenta el rendimiento del potrero al ser regular. Las reglas que serán activadas son las del grupo de 3 antecedentes, en este caso la variable del potrero se activa porque si afecta de manera negativa el engorde animal.

IV-B3b. Los datos de entrada son los siguientes:

Cuadro IX: Representación de datos del ganado y clasificación del estado Animal en el experimento 3.

Edad (meses)	Peso inicial(kg)	Peso fin(kg)	ID animal	Predicción	Grados pertenencia
21	434.8377	438.0388	V1	0	0.5201
29	450.8377	456.0388	V2	1	0.5201
23	429.8377	435.0388	V3	1	0.5201
22	463.8377	469.0388	V4	1	0.5201
22	446.8377	452.0388	V5	1	0.5201
22	465.8377	471.0388	V6	1	0.5201
22	430.8377	436.0388	V7	1	0.5201
27	443.8223	448.0238	V11	0	0.4201
23	451.8223	456.0238	V12	0	0.4201
27	448.8223	453.0238	V13	0	0.4201

El cuadro IX presenta una vista parcial de los datos históricos de engorde del ganado. Basado en lo dicho para los escenarios anteriores, notamos que hay varios animales que están fuera del rango ideal (ver primeras 4 columnas). Se busca que el sistema clasificador difuso los identifique, ahora usando también la variable del potrero.

Usando nuestro sistema clasificador, se ve en el cuadro IX, en la columna grado de pertenencia a la etiqueta estable en bajo el algunos casos (el determina ese valor). Esto quiere decir que pertenecen muy poco a la etiqueta estable (por ello se observa en la columna predicción su etiqueta con el valor 0).

Cuadro X: Datos del potrero del experimento 3.

Nombre	Clima	Potrero
BRS Zuri	Seca	1

El cuadro X contiene los datos utilizados para la entrada del potrero. Se muestra el potrero ocupado, el nombre del forraje usado en el potrero y la condición climática.

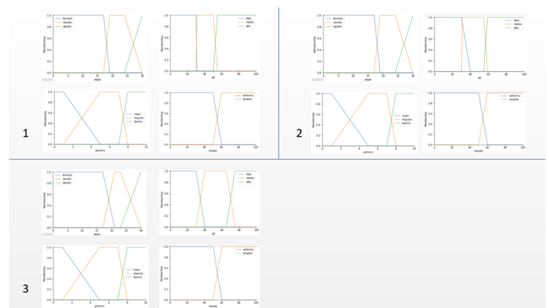


Figura 16: Adaptación de las funciones de pertenencia.

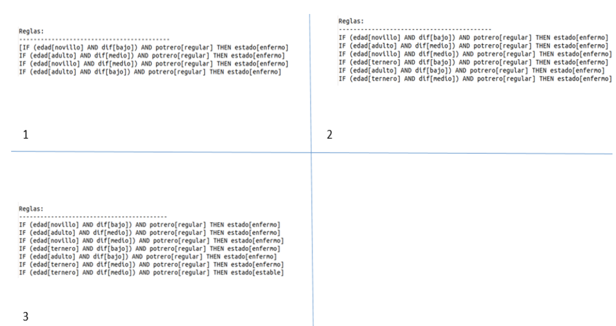


Figura 17: Evolución de las reglas que se activan en el caso 3.

Se observa en la Figura 16 las 4 variables difusas (ahora con el potrero), y como se van ajustando sus conjuntos difusos. En la Figura 16.1 se ve el mejor individuo de la primera iteración y las reglas activadas con el dataset de entrenamiento en la Figura 17.2, y así sucesivamente a través de las iteraciones. En la Figura 16.3 se muestra la solución final del AG cuando converge (individuo con el mayor valor en la función de aptitud), y en la Figura 17.3 se muestra las reglas activadas con el dataset de entrenamiento. Este será la base de reglas difusas final. Hay un número mayor de reglas, ya que se necesitan más reglas que controlen el sistema.

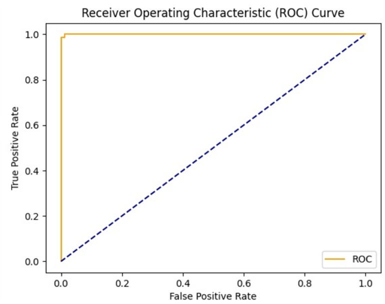


Figura 18: Curva ROC en el caso 3.

Cuadro XI: Métricas del caso 3.

Métricas	Valores
Precisión	95 %
Certeza de reglas	R1:0.88, R2:0.75, R3:0.89, R4:0.85, R5:0.85, R6:0.75
AUC	1

Finalmente, con las reglas difusas finales se procede a probar el sistema usando los datos de prueba, y se calculan las métricas de calidad (ver Tabla XI y Figura 18).

V. Comparación con otros trabajos

Para comparar este trabajo con otros similares, se procedió a definir cuatro criterios, los cuales son:

- $Cri1$: Usa esquemas no-intrusivos para el diagnóstico.
- $Cri2$: Los trabajos usan lógica difusa en el modelo de diagnóstico.
- $Cri3$: Los trabajos usan aprendizaje de máquinas para mejorar el modelo de diagnóstico.
- $Cri4$: Los trabajos analizan conjuntamente el pastoreo y el bienestar animal.

Cuadro XII: Comparación con otros trabajos.

	Cri1	Cri2	Cri3	Cri4
[2]	✓	✗	✓	✗
[3]	✓	✗	✓	✗
[7]	✗	✓	✓	✗
[5]	✗	✓	✓	✗
Este trabajo	✓	✓	✓	✓

Estos criterios son relevantes en la ganadería de precisión porque cumplen 2 aspectos importantes, que son el uso de tecnologías automáticas de la industria 4.0, y procura el bienestar animal para mejorar la producción ganadera. A continuación, se muestra en el Cuadro XII la comparación con previos trabajos.

El primer criterio lo cumplen [2] y [3], ya que ambos usan métodos y modelos computarizados para su diagnóstico. El segundo criterio lo cumplen [7] y [5], ya que usan lógica difusa en su arquitectura, en un caso para encontrar enfermedades malignas, y en el otro para la generación de un sistema de producción medicinal. El tercer criterio lo cumplen todos, ya que usan el aprendizaje de máquinas para diferentes cosas, por ejemplo, para procesar la información acústica del mascado animal [2], estudiar el entorno del ganado [3], detectar enfermedades [7], o establecer el mejor sistema de producción medicinal [5]. La cuarta característica solo la cumple nuestro trabajo, ya que está dirigido a la ganadería de precisión en un entorno de pastoreo rotacional que busca el bienestar animal. Este trabajo cumple con todos los criterios que apuntan a un modelo de diagnóstico para ganadería de precisión. Asegura el bienestar animal a través de un clasificador difuso, usa un modelo de diagnóstico adaptativo basado en aprendizaje evolutivo (usando AGs), y es usado en ganadería precisión para asegurar la salud animal y sacar el mejor provecho al ganado en pastoreo rotacional.

VI. Conclusiones

La presente investigación presentó el desarrollo de un sistema clasificador difuso que supervisa el proceso de engorde animal en potreros rotacionales. El clasificador fue diseñado integrando FCM, un razonador difuso y AGs, para obtener así un modelo de diagnóstico adaptativo a los datos del entorno. El sistema clasificador difuso propuesto permite analizar los datos para proponer un conjunto de reglas que permita diagnosticar el proceso de engorde de un lote de ganado, para lo cual, posteriormente, usa un sistema de inferencia difusa. Este sistema clasificador se caracteriza por su flexibilidad y su tolerancia a la imprecisión [20], ya que puede realizar un razonamiento aproximado usando información del entorno (los datos de entrada al sistema clasificador). Además, nuestro sistema aprende las reglas para el diagnóstico del proceso de engorde del ganado.

La efectividad del método propuesto se ha demostrado mediante varios casos experimentales, con resultados que son prometedores. Las métricas de precisión tienen un alto valor, lo cual indica que el sistema de diagnóstico logra entrenarse bastante bien usando los datos del proceso, con una baja taza de error en cuanto a falsos positivos. Por otro lado, las métricas de certeza son superiores a 0.75, lo que indica que los niveles de activación de las reglas actualizadas son altos, ya que se adaptan a los datos (definen la utilidad de esas reglas). Finalmente, AUC es cercano a 1, lo que dice que el clasificador tiene un margen de error muy bajo, y su nivel de equivocación es casi nulo. Todos estos valores confirman una credibilidad alta en el diagnóstico del sistema.

Esta propuesta es una alternativa efectiva para ser aplicada en ganadería de precisión, específicamente en el pastoreo rotacional para diagnosticar enfermedades del ganado, ya sea por enfermedad o por el rendimiento del potrero. Sin embargo, entre las limitaciones de este trabajo están que el sistema es solo aplicable a pastoreo rotacional, la población de ganado debe ser del trópico, y se evalúan solo 2 estaciones climáticas (verano e invierno). Así, este sistema solo puede ser utilizado en contextos con las siguientes características: en fincas que utilicen la técnica de pastoreo rotacional, sin importar el tamaño de los lotes de ganado, y que las estaciones climáticas sean verano e invierno, entendiendo que en el trópico el invierno son temporadas de lluvia y verano de sequía.

Uno de los trabajos futuros es incorporar este sistema de diagnóstico en un ciclo autónomo de tareas de análisis de datos [21] para la supervisión del proceso de engorde animal. Este ciclo permitiría automatizar la vigilancia del proceso de engorde animal en el marco de la ganadería de precisión. Otro trabajo futuro es extender el proceso de adaptación de las reglas (que actualmente es solo basado en el ajuste a las funciones de membresías del tipo trapezoidal). Por ejemplo, permitir la adaptación de los conjuntos difusos con otras funciones de membresías (por ejemplo, gaussianas), o incluso, la posibilidad de hacer cambios en las variables usadas en el antecedente de las reglas. En esa misma línea, otras posibles extensiones es usar más variables del contexto (por ejemplo, explícitamente el clima en los antecedentes), permitir que varié el número de conjuntos difusos en las variables difusas (por ejemplo, más o menos estados para caracterizar el proceso de engorde), entre otras mejoras.

Acknowledgment

Todos los autores desean agradecer a la Vicerrectoría de Descubrimiento y Creación, de la Universidad EAFIT, su apoyo en esta investigación. Asimismo, los autores quieren agradecer al proyecto HAMADI 4.0, código 22-STIC-06, del programa regional STIC-AmSud, por el apoyo a este trabajo.

Referencias

[1] F. E. Palominos, H. M. Díaz, S. K. Palominos, and L. R. Cañete, “Relación entre los procedimientos de selección a la educación superior y el desempeño académico de los estudiantes con base en una clasificación mediante conjuntos difusos,” *Formación universitaria*, vol. 11, no. 1, pp. 45–52, 2018.

[2] J. O. Chelotti, *Sistema de adquisición y análisis de información acústica para ganadería de precisión*. PhD thesis, 2018.

[3] Y. Qiao, M. Truman, and S. Sukkarieh, “Cattle segmentation and contour extraction based on mask r-cnn for precision livestock farming,” *Computers and Electronics in Agriculture*, vol. 165, p. 104958, 2019.

[4] R. García, J. Aguilar, M. Toro, and M. Jiménez, “Weight-identification model of cattle using machine-learning techniques for anomaly detection,” in *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 01–07, 2021.

[5] J. A. Ramírez Urrego, H. O. Sarmiento Maldonado, and J. M. López Lezama, “Diseño de un clasificador difuso para el establecimiento de los estados funcionales de un sistema de producción de aire medicinal,” 2017.

[6] M. León, B. L. de Lameda, C. Lameda, J. G. Chacon, M. S. Martínez, J. Rojas, J. Contreras-Velásquez, M. Graterol-Rivas, S. Wilches Durán, M. Aguirre, *et al.*, “Aplicación de lógica difusa y algoritmos genéticos para clasificación de tratamientos contra enfermedades neoplásicas malignas,” *Archivos Venezolanos de Farmacología y Terapéutica*, vol. 35, no. 2, pp. 36–41, 2016.

[7] G. Gabriel Elías Chanchí, M. Luz Marina Sierra, and M. Wilmar Yesid Campo, “Application of fuzzy logic in the implementation of evaluation rubrics in the university context [aplicación de la lógica difusa en la implementación de rúbricas de evaluación en el contexto universitario],” *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, vol. 2021, no. E42, pp. 164–177, 2021.

[8] M. Cerrada, J. Aguilar, E. Colina, and A. Titli, “Dynamical membership functions: an approach for adaptive fuzzy modelling,” *Fuzzy Sets and Systems*, vol. 152, no. 3, pp. 513–533, 2005.

[9] J. Aguilar, “A fuzzy cognitive map based on the random neural model,” in *Engineering of Intelligent Systems* (L. Monostori, J. Váncza, and M. Ali, eds.), (Berlin, Heidelberg), pp. 333–338, Springer Berlin Heidelberg, 2001.

[10] H. Sánchez, J. Aguilar, O. Terán, and J. Gutiérrez de Mesa, “Modeling the process of shaping the public opinion through multilevel fuzzy cognitive maps,” *Applied Soft Computing*, vol. 85, p. 105756, 2019.

[11] A. C. Martínez and A. D. A. Armenteros, “La lógica difusa en la modelización del riesgo operacional. una solución desde la inteligencia artificial en la banca cubana fuzzy logic to model operational risk: A solution based on artificial intelligence in the cuban banking,” *COFINHABANA*, no. 2, pp. 122–135, 2018.

[12] M. Cerrada and J. Aguilar, “A fuzzy classifier system for fault tolerance,” *Revista Técnica de la Facultad de Ingeniería. Universidad del Zulia*, vol. 23, 1 1.

[13] C. A. Riveros Varela, M. A. Melgarejo Rey, A. Riveros Varela, and L. D. Alvarado Nieto, “Sistema difuso evolutivo para la predicción del nivel de contaminación del aire por material particulado: Caso puente aranda (bogotá),”

[14] R. García, J. Aguilar, M. Toro, A. Pinto, and P. Rodríguez, “A systematic literature review on the use of machine learning in precision livestock farming,” *Computers and Electronics in Agriculture*, vol. 179, p. 105826, 2020.

[15] P. Melin, F. Olivas, O. Castillo, F. Valdez, J. Soria, and M. Valdez, “Optimal design of fuzzy classification systems using pso with dynamic parameter adaptation through fuzzy logic,” *Expert Systems with Applications*, vol. 40, no. 8, pp. 3196–3206, 2013.

[16] I. S. Masich, M. A. Kulachenko, P. S. Stanimirović, A. M. Popov, E. M. Tovbis, A. A. Stupina, and L. A. Kazakovtsev, “Formation of fuzzy patterns in logical analysis of data using a multi-criteria genetic algorithm,” *Symmetry*, vol. 14, no. 3, p. 600, 2022.

[17] A. K. Dubey, U. Gupta, and S. Jain, “Comparative study of k-means and fuzzy c-means algorithms on the breast cancer data,” *International Journal on Advanced Science, Engineering and Information Technology*, vol. 8, no. 1, pp. 18–29, 2018.

[18] J. Cózar, J. A. Gámez, *et al.*, “Learning compact zero-order tsk fuzzy rule-based systems for high-dimensional problems using an apriori+ local search approach,” *Information Sciences*, vol. 433, pp. 1–16, 2018.

[19] C. Ramos-Palencia, D. Mújica-Vargas, and M. Mejía-Lavalle, “Análisis comparativo de la calidad de agrupamiento del algoritmo intuitionistic fuzzy c-means.,” *Res. Comput. Sci.*, vol. 149, no. 8, pp. 365–378, 2020.

[20] J. Cabrera, “Classical logic and diffuse logic: facets that characterize them,” *Estoa*, vol. 2, pp. 91–101, 05 2013.

[21] J. Aguilar, O. Buendia, A. Pinto, and J. Gutiérrez, “Social learning analytics for determining learning styles in a smart classroom,” *Interactive Learning Environments*, vol. 30, no. 2, pp. 245–261, 2022.

Appendix E

A multi-objective optimization model to maximize cattle weight-gain in rotational grazing

A multi-objective optimization model to maximize cattle weight-gain in rotational grazing

Rodrigo García[1,2,3], Marvin Jiménez[1], Jose Aguilar[1,4,5],
[1]GIDITIC, Universidad EAFIT, Medellin, Colombia.
[2]Faculty of Engineering, Universidad de Córdoba, Montería, Colombia.
[3]Faculty of Engineering, Universidad del Sinú, Montería, Colombia.
[4]CEMISID, Universidad de Los Andes, Merida, Venezuela.
[5]IMDEA Networks Institute, Leganés, Madrid, Spain.
rjgarciah@eafit.edu.co,marvinjimeneznarvaez@gmail.com, aguilar@ula.ve,

Abstract

Rotational grazing can improve cattle feeding considering a series of aspects, such as: (1) the maximum utilization of each hectare of pasture on which the cattle are fed and; (2) the pasture analysis to guarantee the type and the size of the pasture that will serve as feed, among others. The above aspects allow a better-fed cattle, with better weight and meat quality. To implement rotational grazing, it is necessary to carry out forage-utilization practices with criteria associated with the morphophysiology and phenology of the forage species. Many of these data, in the real context of a beef farm, are not used, and most of the decisions made by the farmer are based on experience (successes and failures in productivity). In this proposal, we establish a multi-objective rotational-grazing assignment model based on (1) the best quality forage, and (2) the distance an animal must travel from one paddock to another. At each stage, we estimate the amount of forage as well as the total weight of each batch of animals. Based on pasture yield and cattle forage demand, we propose a dynamic assignment model. To validate the effectiveness of the assignment model, we carried out a discrete simulation of a one-year cattle rotation. Results show that the assignment model generates a statistically higher average weight gain than the one generated by the traditional rotation method.

keywords: Scheduling, Multi-objective Optimization, Artificial Intelligence, Precision Livestock Farming, Rotational Grazing

1 Introduction

Livestock grazing is a predominant type of land use, providing the livelihood for more than a billion people (Quaas & Baumgärtner, 2012). In rotational grazing, a beef farm divides its land into smaller parcels by fencing (electric or wire). Its main objective is to maintain a balance between forage and the nutritional needs of the cattle (Bailey & Brown, 2011). Rotational grazing has been used in agriculture for many years, and has been accepted as a more efficient and sustainable alternative to continuous grazing (Mayee Chen, 2018).

Thus, for the same amount of grass in both situations, rotational grazing can support more cattle and it is thus more productive (Mayee Chen, 2018). In addition to natural factors, overgrazing is the main cause of rangeland ecosystem degradation (Padilla & Sardiñas, 2005). Rotational grazing is a reasonable form of grazing to alleviate overgrazing, increase rangeland productivity, and ecosystem functionality can be improved to some extent. In general, the times of occupancy, rest and paddock assignment to be occupied in rotational grazing are based on the subjective experience of the livestock farmer (Xiaoyan, Zirui, Yingying, Dengsheng, & JianPing, 2019a).

Some papers related to the grazing optimization problem are presented as follows. White *et al.* (White, Brady, Capper, & Johnson, 2014) developed a model that optimizes pasture and nutritional management to examine the environmental impact of beef production. White *et al.*'s model integrated modules that calculate (1) environmental impact from cradle to farm gate, (2) diet cost, (3) pasture growth and (4) willingness to pay. Xiaoyan *et al.* (Xiaoyan, Zirui, Yingying, Dengsheng, & JianPing, 2019b) established a dynamic model of rotational grazing assignment, which combines the assignment model and the carrying capacity constraint, as a function of the forage yield of the rotational grazing plots and the forage demand of the cattle. Chen *et al.* (Chen & Shi, 2018) constructed a differential-equation model of grazing with vegetation on a fixed area, to show that production yields and stored forage are higher in rotational grazing than in continuous grazing; their results showed that both head of cattle per hectare and stored forage increased for many rotational configurations. Raizada *et al.* (Raizada, Dogra, & Dhyani, 2008) used multi-objectives to develop alternative land use plans to optimize four objective functions: maximizing (1) farm income, (2) employment and (3) nutritional security, and minimizing (1) soil loss ii) individually at the watershed level, ii) annual forage production, to guarantee sustainable animal population and changes in forest dependence. Qin *et al.* (Qin et al., 2018) developed an optimization model for yak grazing in alpine meadows. They solved the optimization model of multi-objective alpine meadow grazing by means of a genetic algorithm. Also, Du et al. (DU et al., 2018) propose a model to optimize the yak population using an evolutionary algorithm, taking the maximum net monetary income as the objective function. The proposal defines an optimal scheme for selling yaks without destroying the ecological environment, and the experimental results show the validity and feasibility of the model. Liang *et al.* (Liang, Feng, Xia, & Cui, 2011) used a multi-objective approach in Matlab to define an optimization and management strategy for maintaining a balance between forage supply and livestock requirement, based on dynamic monitoring data from rangeland livestock husbandry. Addis *et al.* (Addis, Blair, Kenyon, Morris, & Schreurs, 2021) developed a linear programming profit optimization model with a silage supplementation scenario. The usable kilograms of pasture dry matter from total pasture mass were derived using the minimum and maximum pasture mass available to cattle and sheep and percent grass utilization. They employed linear programming to identify the optimum carrying capacity of cattle and sheep, the most profitable slaughter ages of cattle, the number of prime lambs (sold to meat processing plants), and the reserve lambs sold (sold to other farmers for finishing).

All previous works focused on grazing optimization from different approaches; however, none focused on selecting the highest quality forage, maximizing weight gain of bovine animals, and the distances the animal has to travel from one paddock to another. These limitations were highlighted in a recent systematic literature review (García, Aguilar, Toro, Pinto, & Rodríguez, 2020). To overcome these limitations found in previous works, this article focuses on the use of variables of beef production for optimal grazing decisions; an approach towards autonomous or semi-autonomous beef production.

This paper presents a rotational-grazing assignment model based on (1) the best quality forage and (2) the distance an animal has to travel from one paddock to another. First, in each assignment stage, we estimate the amount of forage per paddock (differentiating the canopy from the rest of the forage), as well as the total weight of each batch of animals, using different data capture methods (e.g., drones, cameras and sensors). Then, based on paddock performance and cattle forage demand, a dynamic assignment model is established. Each day, variables such as carrying capacity, quality forage and distances between paddocks are evaluated, and an assignment of cattle lots to paddocks is made, maximizing the total weight gain of the lots and minimizing the distance an animal must travel from one paddock to the next. Thus, a feasible solution for a rotational grazing cycle is obtained. The hypothesis is that a reasonable dynamic rotational-grazing decision scheme can make resources efficient in time and space. The main contributions of this work are:

- Determination of forage quality and its effect on animal weight gain.
- Minimization of the distance moved by a herd to a paddock
- Definition of a dynamic approach for this assignment problem
- Execution of a Multiple herd assignment

This work is organized as follows. Section 2 introduces the assignment mathematical model used in this work. Section 3 shows our approach through different case analyses in meat production. After, Section 4 compares this work with previous work. Finally, Section 5 presents the conclusions and future works.

2 Assignment mathematical model

In this section, we present the problem, and the formulation of the optimization model.

2.1 Assignment problem of rotational grazing

An assignment problem consists of the allocation of resources to perform tasks, with the objective of satisfying desired goals (e.g., maximizing benefits or minimizing costs) (Bikhchandani & Ostroy, 2002; Aguilar, 1998, 2001).

Rotational grazing consists of dividing the entire area of a farm into more than two paddocks, while some remain occupied, the others are at rest (Briske et al., 2008) (see Figure 1). This reduces the total grazing area, and forces cattle to consume forage uniformly, assigning lots of cattle to different paddocks (Jacobo, Rodríguez, Bartoloni, & Deregibus, 2006). So, this problem can be treated as an assignment problem. The relationship between the resource and the task in the assignment model is equivalent to the corresponding relationship between herds and paddocks.

Rotational grazing and continuous grazing are two different methods of managing pasture and livestock on a farm. Rotational grazing involves dividing the property into different fields or plots, and moving livestock from one field to another in a prearranged order. This allows the grass to have time to recover in each field before being grazed again. Continuous grazing, on the other hand, involves keeping cattle in a single field and allowing them to graze freely throughout the property (da Silva, Imhoff, & Corsi, 2003). One of the main differences between these two methods is the impact they have on pasture health. In rotational grazing, the grass has time to recover between grazing, allowing it to grow healthier and more sustainably. This is because the grass is not grazed too often, which allows it to maintain a high amount of chlorophyll and good root health. On the other hand, in continuous grazing, the grass may be grazed too often, which can result in a lower amount of chlorophyll and a weaker root (Derry & Boone, 2010).

In a rotational-grazing cycle, it is necessary to distribute the herd in different paddocks. Roughly speaking, this process consists of:

1. Prepare the paddocks by sowing the pasture
2. Buy the herds of animals
3. Assign an initial paddock to each herd
4. Each day x, assign a new paddock to each herd
5. Regularly weigh the herds to measure their fattening.
6. Sell the cattle when they reach a given weight

Some of the variables that can be considered in this process are:

- Pasture: source of food for livestock.
- Climate: the average weather conditions that characterize a given place.
- Forage capacity: the amount of forage in each paddock.
- Residence period: is the total time in hours or days that an animal or group of animals grazes in each rotation.
- Occupancy period: The total time spent grazing a paddock by all groups of animals in each rotation; when there is only one group of animals, the period of stay is equal to the period of occupancy.
- Resting or recovery period: The period between two successive grazings, during which the pasture is allowed to rest.

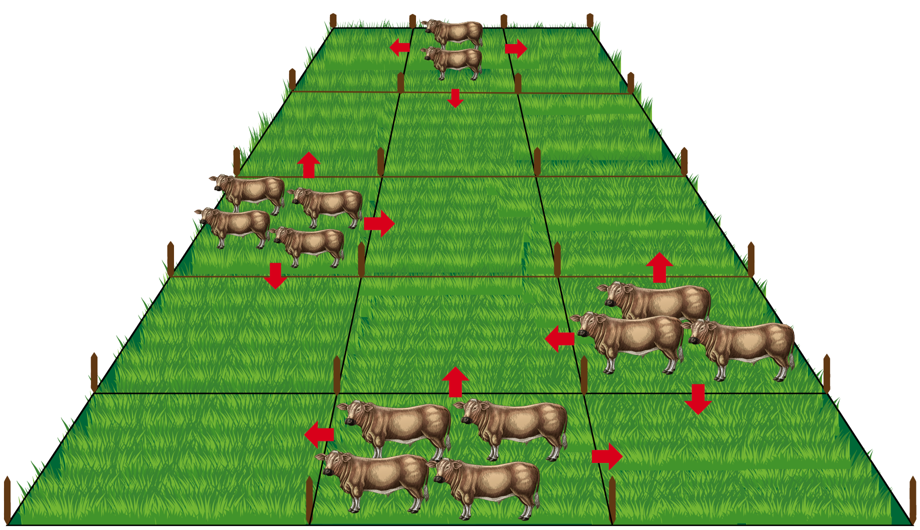


Figure 1: A rotational-grazing system

2.2 Assignment model

In the process of herd rotation in paddocks, quality consumable forage must be considered. In this process, it is necessary to ensure the nutritional demand of the herd and the pasture resources (Romanzini et al., 2022). Therefore, this paper chooses the maximum total yield of quality forage as the optimal situation in the grazing cycle.

The optimization proposed in this work is focused on the allocation of lots to paddocks, for which we seek to make an optimal allocation considering criteria such as weight gain, distance traveled by the lots and the use of quality forage. This work seeks to propose an allocation model that improves the process of lot rotation, but not an optimization within a paddock, in which the animals are not controlled and are left to graze at their own free will. The idea is to optimize the process that can be controlled, which is the allocation of lots to paddocks.

To understand the mathematical model, tables with the notation used are presented first, starting with the indices and the decision variable in Tables 1 and 2.

Table 1: Indexes in the assignment model.

Indexes	Description
i	Index of herds
j	Index of paddocks
t	Time index (days)

Table 2: Decision variable in the assignment model

Decision Variable	Description
x_{ij}^t	It is 1 if the herd i is assigned to the paddock j at time t; otherwise, it is 0.

Table 3 shows the definitions of the model parameters. These parameters correspond to the state of the grazing system at the time that the assignment model is run.

Table 3: Parameters of the assignment model.

Parameters	Description
w_j^t	Estimated forage quality at a time t.
p_i^t	Percentage of nutritional need of herd i estimated at a time t
q_i^t	Total weight in kilograms of herd i estimated at a time t
g_j^e	Average daily weight gain in kilograms on paddock j at a time e of year.
C_i^t	Number of animals in the herd i.
a_i	Total area occupied by herd i.
A_j	Total area of paddock j.

Table 4 shows the definitions of the auxiliary variables of the assignment model.

Table 4: Auxiliary variables of the assignment model.

Variables	Description
d_{ij}^t	Occupation time of herd i consuming quality forage from paddock j, estimated at time t.
G_{ij}^t	Weight gain to be obtained by herd i in paddock j estimated at time t.
D_{ij}^t	Distance in meters between herd i and paddock j at a time t.

With the nomenclature described above, the mathematical model with all its equations is presented below. The assignment model has two objectives: to maximize the weight gain of the animals and to minimize the distance traveled when they are moved from one paddock to another. The first objective is represented by equation 1, which maximizes the sum of the weight gains that the animals would obtain if they were assigned to each of the paddocks at time t. On the other hand, the second objective (see equation 2) is achieved by minimizing the sum of the total distance traveled by the animals when they are assigned to a paddock at a time t.

$$Max\ \ Z_1 = \sum_{i=1}^{n}\sum_{j=1}^{m} G_{ij}^t x_{ij}^t \tag{1}$$

$$Min\ \ Z_2 = \sum_{i=1}^{n}\sum_{j=1}^{m} D_{ij}^t x_{ij}^t \tag{2}$$

S.t.

The optimization process must comply with the following restrictions.

Constraint 1: Since only one lot can be in each paddock at a time, equation 3 is included as the first constraint of the model. This restriction ensures that at each time t all paddocks that are occupied are only occupied by a single lot.

$$\sum_{i=1}^{n} x_{ij}^t = 1,\ \ \forall j \tag{3}$$

Constraint 2: In the event that the number of lots is less than the number of paddocks, the model creates fictitious lots to achieve equality so that the model always satisfies the condition that $n = m$. At the moment of implementation in real life the optimal allocation obtained by solving the model, the paddocks where a fictitious lot is assigned correspond to empty paddocks.

Constraint 3: Similarly, all lots are assigned to some paddock, and it is unique, i.e., a lot cannot be assigned to more than one paddock (see equation 4).

$$\sum_{i=1}^{m} x_{ij}^t = 1,\ \ \forall j \tag{4}$$

Constraint 4: When considering the allocation of a lot i in any paddock, the space occupied by the lot and the area of the paddock under consideration must be taken into account, such that the area occupied by the lot must not exceed the area of the paddock (see Equation 5).

$$a_i x_{ij}^t \leq A_j, \ \ \forall j \tag{5}$$

Constraint 5: To calculate the total weight gain that a flock would obtain if assigned to a specific paddock, it is necessary to calculate the time (in days) that the flock would remain occupying the paddock until consuming all the quality forage in the paddock. Additionally, the average daily gain of an animal is included, which depends on the time of the year and the number of animals in the flock. This total weight gain is calculated by means of equation 6.

$$G_{ij}^t = d_{ij}^t g_j^e C_i^t, \ \ \forall i, \forall j \tag{6}$$

Constraint 6: The occupancy time of a given lot in a specific paddock is calculated by dividing the amount of quality forage in the paddock by the amount of grass the lot needs in a day (see Equation 7).

$$d_{ij}^t = \frac{w_{ij}^t}{p_i^t q_i^t}, \ \ \forall i, \forall j \tag{7}$$

Constraint 7: Finally, this last restriction corresponds to the binary nature of the decision variables (Equation 8).

$$x_{ij}^t \in \{0, 1\}, \ \ \forall i, \forall j, \forall t \tag{8}$$

The proposed model is a dynamic multiobjective optimization model. It is dynamic optimization because the model takes into account the evolution of variables and parameters over time. The model is run every day after updating information related to paddocks and animals (forage, weight, etc.). Each day t the best possible allocation of animal lots to paddocks is made, including the possibility that some lots should not change paddocks. To perform the allocation, the model makes an estimation of the weight gain that each batch of cattle would obtain in each possible paddock over the considered time window. Additionally, the model is multi-objective, it seeks to maximize the weight gain of the animals and minimize the distance they travel when they are moved to another paddock.

The optimization is done by solving the mixed integer linear programming mathematical model (objective function and constraints) using a library of R. The model is solved each day during the simulation, i.e., each day the optimal allocation of lots to paddocks is obtained, and the time spent in the lots is recalculated.

3 Model analysis

For the analysis of the model, we present the simulator, the experiments and the result analysis.

3.1 Simulator of a cattle rotation system

To validate the effectiveness of the assignment model, we carried out a discrete simulation of one-year of a cattle rotation system. To adequately represent the real dynamics of beef production, we considered (1) the effect of the time of the year (rain, drought) on the grass-growth rate, (2) plant species and (3) its flowering time. Changes in system states were analyzed daily.

The number of paddocks and animal flocks are parameters of the simulator. For each paddock, characteristics –such as area, location within the farm, type of plant and capacity– are randomly generated. Likewise, for each herd, the number of animals that make up the herd, as well as the weight, age, and genre of each one of them, are randomly produced.

At the beginning of the simulation, all the paddocks have a complete forage defined as (1) quality forage and (2) non-quality forage. The total forage of each paddock changes every day due to the consumption of the herd that occupies it, or its natural growth or decrease caused by rain or drought. The quality forage is found in the upper part of the pasture and is consumed first by the animals because of its flavor and because it is the first part of the plant with which they come into contact. For this reason, the simulation assumes that the quality forage of a paddock is the first to be consumed, and after it is exhausted, the consumption of the lower-quality forage begins. The weight of the animals evolves considering (1) the quantity and quality of forage consumed, (2) age, (3) genre, and (3) distances traveled when the herd moves to another paddock.

The number of simulated paddocks is greater than the number of herds. The herds are randomly assigned to their respective paddocks at the beginning of the simulation, then they are moved to other paddocks according to some criteria, such as (1) availability of forage in the paddocks and their distance and (2) the nutritional needs of the herd.

The simulator needs several parameters to define the context. The parameters related to the *paddocks* are: (1) the daily growth rate of the pasture (in percentage units), (2) the plant species, (3) the rate of extra increase in the rainy season, (4) the rate of loss in the dry season, (5) the rate of loss due to flowering, (6) the daily weight gain of an animal depending on the quality of the forage, (7) the minimum and maximum area of a paddock, (8) the minimum and maximum capacity of the paddocks at the beginning of the simulation, (9) the measurements of the farm within which the paddocks are randomly located before starting the simulation, (10) the initial fraction of the total forage that is quality forage and (11) the amount of forage per square meter that grows in a paddock on a rest day after the capacity reaches zero, (12) the number of paddocks, (13) the area of each paddock (m^2), (14) location of the paddocks within the farm, (15) forage of each paddock (kg), (16) number of consecutive days of occupation allowed per paddock, and (17) number of days that a paddock must remain unoccupied after the maximum number of consecutive days of occupation allowed. These last two parameters are used also in the traditional grazing simulation scenario (without optimization model).

In the *case of herds*, the parameters are (1) the number of animal lots, (2) the nutritional requirement, as a percentage of the total weight of a cattle lot that the lot needs to consume daily to increment the weight, (3) the minimum and maximum number of animals per lot, and (4) the weight loss per walk (kg/m), (5) the daily nutritional requirement of an animal (% of its weight), Finally, the last parameter is the number of days to be simulated.

the *output variables* are: (1) Final weight of animals, (2) Average weight of animals, (3) Average weight gain of the animals, (4) Final forage of each paddock (quality and non-quality) (5)Average forage (quality and non-quality)

In summary, the discrete event simulator macro-algorithm of a cattle rotation system is:

1. Define the values of the simulator parameters

2. Generate randomly weather data from real information.

3. Obtain the paddock distance matrix.

4. Then, for each day t of the time window to be simulated:

(a) Read the weather characteristics, and for each pasture type in the paddocks check if it is the flowering season.

(b) Perform the assignment of lots to paddocks (either traditionally or using the optimization model).

(c) Update the weight of the animals in each batch of cattle by calculating the weight gain obtained in the respective paddocks to which the lots were assigned.

(d) Update the forage of each paddock considering the consumption of grass by the assigned batch of animals, the daily growth of the grass, the increase due to rainfall, or the loss due to drought or flowering.

This process is carried out using functions for the calculation of animal weight gain, grass increase due to rainfall, forage loss due to drought, forage loss due to flowering season, allocation of lots to paddocks and daily forage growth. The simulator is located at https://github.com/SimuladorGanadero

On the other hand, in the literature, there are tables that indicate the daily weight gain obtained by an animal depending on the type of pasture consumed. In the simulation, these values were used as parameters, and were increased or decreased depending on the time of the year (rain, drought). The increase or decrease factor is a parameter of the simulation. Finally, the grass growth rate is calculated during the simulation taking into account the grass species and climate, which are inputs of the simulation.

3.2 Experimental Scenarios

To evaluate the performance of the mathematical model, two strategies were used with the same values for the parameters described above in the simulator, but differing in the herd rotation dynamics in the paddocks. The simulation setup performed for the experimentation is presented in Table 5:

Table 5: Simulation parameters

Parameter	**Value**
Days of simulation	365
Number of batches	15
Number of animals per flock	5
Number of paddocks	100
Minimum area of a paddock (m^2)	6000
Maximum area of a paddock (m^2)	7000
Minimum capacity of a paddock (kg of grass)	3000
Maximum capacity of a paddock (kg of grass)	3500
Maximum number of consecutive days a paddock can be occupied consecutively.	3
Ideal number of days a paddock should remain unoccupied after being used.	25
Forage (kg/m^2) that grows in one day in a paddock after it has been completely consumed.	0.08
Fraction of total forage that is quality forage	0.3
Weight loss per walk (kg/m)	0.00001
Fraction of weight gain that is in addition to the average gain for quality forage	0.15

Table 5: Simulation parameters

Parameter	Value
Daily nutritional requirement of an animal (percent of its weight)	11%
Prime rate of daily growth of grass (forage)	12%
Increase in forage due to rainfall gain	12%
Decrease in forage due to drought loss	4%
Decrease in forage due to flowering loss	3%

The first strategy: Traditional rotational grazing, in which herds are moved from one paddock to another periodically, considering estimates of forage in the unoccupied paddocks, the distance the herd would have to travel to each possible paddock, and the days that the paddocks must remain at rest due to their capacity reached zero or they were occupied for a certain number of consecutive days. The latter and the rotation period are given in days, and are parameters of the simulation.

The second strategy: The rotation using the proposed optimization model, which is run every day and finds the optimal assignment of the herds to the paddocks, considering the quantity and quality of forage in the paddocks at that moment. In this way, the model determines for each batch of animals whether they should stay in the current paddock, or be moved to another paddock to maximize their weight gain and minimize the distance to be traveled. The proposed model is multi-objective, however, in the simulations performed it is approached as a mono-objective problem by including a factor that quantifies the effect of walking on the weight of the cattle, as shown in equation 9.

$$Max\ \ Z = \sum_{i=1}^{n}\sum_{j=1}^{m}(G_{ij}^{t} - W_l D_{ij}^{t})x_{ij}^{t} \tag{9}$$

where W_l is the amount of weight in kilograms that an animal loses per meter traveled.

3.3 Statistical analysis

For each scenario, 30 simulation runs were performed with different random seeds. The main parameters and their respective values when unchanged are as follows: number of herds (15 herds), number of paddocks (100 paddocks), number of animals per herd (5 animals) and simulation time (365 days). The results are analyzed as follows. The comparison metric for the two simulated grazing alternatives is the average animal weight-gain (AWG), which is given by:

$$AWG = \frac{1}{N}\left(\sum_{i=1}^{N} W_{fi} - \sum_{i=1}^{N} W_{0i}\right) \tag{10}$$

where W_{0i} and W_{fi} are the weights of the animal i at the beginning and end of the simulation, respectively, and N is the total number of animals.

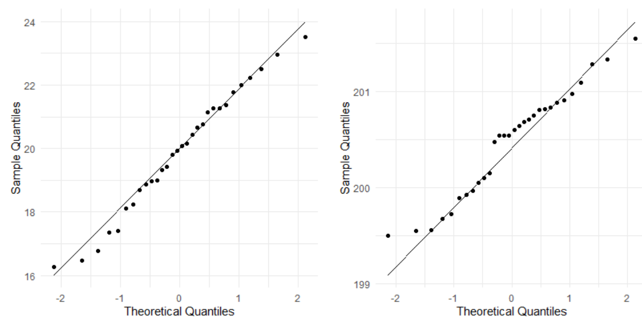
From the simulation of the traditional grazing scenario, a mean average weight gain of 172.94 kg with a standard deviation of 0.11 kg was obtained. For grazing using the assignment model, a mean of 215.52 kgand a standard deviation of 0.089 kg was obtained. Clearly, the average weight gain using the proposed model is superior to that of the using the traditional way. Since these results are random samples, statistical inference is necessary to verify that the proposed model actually produces a higher average weight gain in the animals. Thus, it is desired to test the hypothesis of contrast presented in Table 6.

Table 6: Hypothesis test

Hypothesis	Description of hypothesis
$H_0 : \mu_T = \mu_M$	The means of the AWG are the same.
$H_1 : \mu_T < \mu_M$	The mean AWG with traditional grazing is lower than that of AWG with grazing using the assignment model.

Taking a significance level of 5% ($\alpha = 0.05$), we first check whether the samples resulting from the simulation come from a population with a Normal probability distribution, which is done to know whether it is necessary to apply parametric or non-parametric tests.

Figure 2 shows the quantile-quantile plots comparing the sample quantiles of average animal weight gains with the theoretical quantiles of a Normal distribution. There is a strong positive linear correlation between samples and theoretical quantiles in both samples. This suggests that both samples have a Normal distribution. To verify this hypothesis formally, Shapiro-Wilk Normality test is run, which yields p-values of 0.1901 and 0.9453 for the traditional and model grazing samples, respectively. This indicates that there is no enough evidence to show that the distributions are not normal; thus, we assume that both data sets have a Normal distribution.



(a) Traditional grazing (b) Assignment model

Figure 2: Normal QQ plot

Since the assumption of normality is met, we proceed to verify the equality of variances to identify the appropriate statistical test of means. The F-test(Tian, Manfei, Justin, Hongyue, & Xiaohui, 2018) for comparison of variances of two normal population samples is performed, which yields a p-value of $0.3405 > 0.05 = \alpha$; thus, the population variances are equal.

Therefore, we proceed to perform the T-test (Tian et al., 2018), for comparison of means of two normal populations with equal variances and test the hypotheses, in Table 6. A p-value of $2.2e^{-16} \ll 0.05 = \alpha$ is obtained; therefore, the null hypothesis H_0 is rejected in favor of the alternative hypothesis H_1. Therefore, $\mu_T < \mu_M$.

With these results, it can be affirmed that the assignment model generates a statistically higher average weight gain than the one generated by the traditional rotation method.

3.4 Analysis of the behavior of different parameters in the models

In addition, a basic descriptive statistical analysis is performed to compare the performance of traditional grazing with that of the proposed model by changing some parameters of the simulator. We assume the base values of the parameters indicated in the previous section, and analyze the results after modifying each of them.

Regarding the assignment of *multiple herds*, Table 7 shows the mean and standard deviation of the average weight gain in kilograms of 30 simulations for our model and for the traditional rotation method for different numbers of herds. For example, when the herd number is 5, the rotational grazing simulations using the proposed optimization model have a mean average weight gain of 215.57 kg with a standard deviation of 0.185 kg. On the other hand, the traditional method achieves a mean of 192.8 kg and a deviation of 0.47 kg.

It can be seen that as the number of herds increases, the mean average weight gain decreases. This was an expected result since a higher number of herds of animals means a higher rate of pasture consumption, so the amount of available pasture is depleted more quickly, and this in turn, directly affects the weight gain of the animals.

In each of the scenarios proposed by varying the number of herds, the proposed optimization model achieves a higher average weight gain for the animals than that obtained with the traditional rotation method. Thus, our model presents better performance.

Table 7: Mean and standard deviation of the average weight gain for different numbers of herds.

Number of Herds	Average Weight Gain Our model	Average Weight Gain Traditional
5	215.57 (0.185)	192.8 (0.47)
15	211.45 (0.116)	173.02 (0.07)
25	206.12 (0.094)	170.17 (0.05)

Table 8 shows the results for different *numbers of paddocks*. If the number of paddocks increases, then the mean of average weight gains also increases, and their standard deviations decrease. This behavior is observed in both traditional and optimized grazing methods using the proposed mathematical model. Having a greater number of paddocks allows the grazing system has more possible places with available pasture to move the cattle lots. However, clearly, the proposed model shows better performance than the traditional rotational grazing method for the different numbers of paddocks.

Table 8: Mean and standard deviation of the average weight gain for different numbers of paddocks.

Number of paddocks	Average Weight Gain Our model	Average Weight Gain Traditional
100	208.71 (0.099)	173 (0.09)
125	210.26 (0.076)	185.29 (0.062)
150	215.6 (0.072)	193.87 (0.07)

The results for different *numbers of animals* per herd are presented in Table 9. In the case of the traditional grazing method, it is observed that as the number of animals per flock increases, the average weight gain decreases considerably, making this parameter a factor of great influence on the performance of this method.

However, in the case of the optimization model, the average weight gain remains fairly homogeneous despite the increase in the number of animals per flock. The model manages to absorb the variation in the number of animals quite well, being far superior to the traditional method.

Table 9: Mean and standard deviation of the average weight gain for different numbers of animals by lot.

Number of animals by herd	Average Weight Gain Our model	Average Weight Gain Traditional
5	215.52 (0.089)	172.94 (0.11)
10	215.49 (0.079)	68.9 (1.88)
15	215.64 (0.059)	51.82 (0.65)

Table 10 shows the results for different *numbers of days* to be simulated. As expected, the higher the number of grazing days, the higher the average weight gain. But the superiority of the optimization model in average weight gain in the three simulation time scenarios tested is remarkable.

It is important to highlight that as the grazing time increases, the difference between the average weight gain obtained with the model and that obtained with the traditional method is greater. This is due to the fact that as the days go by, the paddocks have less and less pasture available, and it becomes more difficult to make an intelligent allocation through traditional grazing that includes estimations of future behaviors. This is something that our model does quite well, so it adapts much better than the traditional method, and therefore, achieves higher average weight gains.

Table 10: Analysis of the behavior of our model for different numbers of days simulated.

Number of days simulated	Average Weight Gain Our model	Average Weight Gain Traditional
121	55.75 (0.061)	41.9 (0.05)
242	141.58 (0.031)	112.5 (0.09)
365	215.52 (0.119)	173 (0.11)

4 Comparison with Previous Works

In this section, we propose several criteria to compare previous studies related to animal grazing optimization with our approach. In what follows, we make a comparison of this work with previous studies.

- ***Criterion 1:*** the study proposes a mathematical optimization model applied to livestock processes.
- ***Criterion 2:*** the study addresses the rotational grazing problem using an optimization model.
- ***Criterion 3:*** the study proposes to discriminate forage quality and its effect on animal weight gain.
- ***Criterion 4:*** the study considers animal welfare.
- ***Criterion 5:*** the study proposes an optimization.

In Table 11, a qualitative comparison with related studies is made, based on previous criteria.

Table 11: Comparison with previous works.

	Criterion 1	Criterion 2	Criterion 3	Criterion 4	Criterion 5
(White et al., 2014)	✓	✗	✗	✗	✗
(Xiaoyan et al., 2019b)	✓	✓	✗	✗	✗
(Chen & Shi, 2018)	✓	✓	✗	✗	✗
(Raizada et al., 2008)	✓	✓	✗	✗	✗
(Qin et al., 2018)	✓	✗	✗	✗	✗
(DU et al., 2018)	✓	✗	✗	✗	✗
(Liang et al., 2011)	✓	✗	✗	✗	✗
(Addis et al., 2021)	✓	✗	✗	✗	✗
This work	✓	✓	✓	✓	✓

As shown in Table 11, previous studies did not satisfy all the criteria. Specifically, for *criterion 1*, all related research makes use of mathematical optimization models to improve livestock production. For *criterion 2*, Xiaoyan *et al.* (Xiaoyan et al., 2019b) meet this criterion, since they work with rotational grazing by making dynamic assignments. Also, Raizada *et al.* (Raizada et al., 2008) define a multi-objective approach to maximize farm income, employment and nutritional security, and minimize soil loss. Finally, Chen *et al.* (Chen & Shi, 2018) worked with differential equations of grazing with vegetation on a fixed area to show that production yields and stored forage are higher for rotational grazing than for continuous grazing. Our proposal approaches animal grazing as a multi-objective assignment problem.

For *criterion 3*, our proposal is the only one that focused on maximizing weight gain using quality forage, which is taken by our model to calculate the occupancy time of a herd in a paddock. For *criterion 4*, our proposal is the only one that considers animal welfare. It takes into consideration the distance an animal has to move from one paddock to another, which can cause stress to the animal and, additionally, the burning of calories could result in weight loss. Finally, regarding *criterion 5*, Our model is a dynamic optimization model, since the decision variable and parameters change with respect to time. Additionally, the decision at a time t is made considering an estimation of the behavior of the rotational grazing system in a future time window.

5 Conclusions

This paper provides a solution for grazing planning, using a mathematical assignment model for assigning cattle herds to pastures in a cattle rotation system based on (1) the quality forage and (2) the distance the herd has to travel from one paddock to another. In addition, in the model, we consider the assignment of multiple herds. From an environmental standpoint, using the proposed model can alleviate environmental problems caused by overgrazing, and help the farmer avoid under-grazing. For that, using the model, the farmer can know the optimum time of rotation since the allocation model decides the optimal time to move an animal from one paddock to another. Particularly, the allocation model can be used on any livestock farm that uses rotational grazing as a mode of beef production.

Thus, the main contribution of this work is the use of forage quality, and its effect on animal weight gain, and the distance traveled by a herd to a paddock, to determine the allocation of herds to paddocks. In particular, this process is dynamically carried out, and in turn, allowing multiple allocations of herds. The experiments carried out show the versatility of our model for different situations (number of herds, number of paddocks, among others), always obtaining robust results in each of them.

As future work, the five freedoms of animal welfare of the World Organization for Animal Health (Horgan & Gavinelli, 2006) will be considered in an extension of the assignment model: (1) freedom from hunger, thirst and malnutrition, (2) freedom from fear and distress, (3) freedom from physical and thermal discomfort, (4) freedom from pain, injury and disease, (5) freedom to exhibit natural behavior. As another future work, we propose to incorporate cattle-weight (García, Aguilar, Toro, & Jiménez, 2021) into the model to be able to detect anomalies in the fattening process. This would give the allocation model the ability to show which animal is not gaining the weight it should, alerting the farmer to which animal needs to be monitored. Thus, in future work, incorporate this dynamic assignment model into an autonomous cycle of data analysis tasks (Aguilar, Buendia, Pinto, & Gutiérrez, 2022) for monitoring the animal fattening process is natural. The autonomous cycle would enable automation of animal fattening/rotation process monitoring in the framework of precision livestock farming.

Acknowledgements

Rodrigo García is supported by a PhD grant financed by Universidad EAFIT. All the authors would like to thank the "Vicerrectoría de Descubrimiento y Creación", of Universidad EAFIT, for their support on this research. Also, the authors want to thank the HAMADI 4.0 project, code 22-STIC-06, of the STIC-AmSud regional program, for the support of this work.

Conflicts of Interest

The authors declare there are no conflicts of interest.

References

Addis, A. H., Blair, H. T., Kenyon, P. R., Morris, S. T., & Schreurs, N. M. (2021). Optimization of profit for pasture-based beef cattle and sheep farming using linear programming: Model development and evaluation. Agriculture, 11(6), 524.

Aguilar, J. (1998). Definition of an energy function for the random neural to solve optimization problems. Neural Networks, 11(4), 731-737. doi: https://doi.org/10.1016/S0893-6080(98)00020-3

Aguilar, J. (2001, jun.). A general ant colony model to solve combinatorial optimization problems. Revista Colombiana de Computación, 2(1), 7-18.

Aguilar, J., Buendia, O., Pinto, A., & Gutiérrez, J. (2022). Social learning analytics for determining learning styles in a smart classroom. Interactive Learning Environments, 30(2), 245-261. Retrieved from https://doi.org/10.1080/10494820.2019.1651745 doi: 10.1080/10494820.2019.1651745

Bailey, D. W., & Brown, J. R. (2011). Rotational grazing systems and livestock grazing behavior in shrub-dominated semi-arid and arid rangelands. Rangeland Ecology Management, 64(1), 1-9. Retrieved from https://

www.sciencedirect.com/science/article/pii/S1550742411500089 doi: https://doi.org/10.2111/REM-D-09-00184.1

Bikhchandani, S., & Ostroy, J. M. (2002). The package assignment model. Journal of Economic Theory, 107(2), 377-406. Retrieved from https://www.sciencedirect.com/science/article/pii/S0022053101929576 doi: https://doi.org/10.1006/jeth.2001.2957

Briske, D., Derner, J., Brown, J., Fuhlendorf, S., Teague, W., Havstad, K., . . . Willms, W. (2008). Rotational grazing on rangelands: Reconciliation of perception and experimental evidence. Rangeland Ecology Management, 61(1), 3-17. Retrieved from https://www.sciencedirect.com/science/article/pii/S1550742408500078 doi: https://doi.org/10.2111/06-159R.1

Chen, M., & Shi, J. (2018). Effect of rotational grazing on plant and animal production. Mathematical Biosciences & Engineering, 15(2), 393.

da Silva, A. P., Imhoff, S., & Corsi, M. (2003). Evaluation of soil compaction in an irrigated short-duration grazing system. Soil and Tillage Research, 70(1), 83–90.

Derry, J., & Boone, R. (2010). Grazing systems are a result of equilibrium and non-equilibrium dynamics. Journal of Arid Environments, 74(2), 307–309.

DU, M.-y., QIN, X.-f., ZHANG, C., ZHANG, Y.-a., SONG, R.-d., & SUN, Z.-j. (2018). Evolutionary algorithm for solving the problem of grazing yak population in three rivers source region. In 2018 international conference on computer, communications and mechatronics engineering (pp. 624–629).

García, R., Aguilar, J., Toro, M., & Jiménez, M. (2021). Weight-identification model of cattle using machine-learning techniques for anomaly detection. In 2021 ieee symposium series on computational intelligence (ssci) (pp. 01–07).

García, R., Aguilar, J., Toro, M., Pinto, A., & Rodríguez, P. (2020). A systematic literature review on the use of machine learning in precision livestock farming. Computers and Electronics in Agriculture, 179, 105826. Retrieved from http://www.sciencedirect.com/science/article/pii/S0168169920317099 doi: https://doi.org/10.1016/j.compag.2020.105826

Horgan, R., & Gavinelli, A. (2006). The expanding role of animal welfare within eu legislation and beyond. Livestock Science, 103(3), 303–307.

Jacobo, E. J., Rodríguez, A. M., Bartoloni, N., & Deregibus, V. A. (2006). Rotational grazing effects on rangeland vegetation at a farm scale. Rangeland Ecology Management, 59(3), 249-257. Retrieved from https://www.sciencedirect.com/science/article/pii/S1550742406500387 doi: https://doi.org/10.2111/05-129R1.1

Liang, T., Feng, Q., Xia, W., & Cui, X. (2011). Optimization strategy and management decision-making in balancing forage and livestock in gannan pastoral area. Acta Ecol Sin, 31, 1111–1123.

Mayee Chen, J. S. (2018). Effect of rotational grazing on plant and animal production (Vol. 15) (No. 1551-0018$_2018_{23}$93). $Retrievedfrom$ doi: 10.3934/mbe.2018017

Padilla, C., & Sardiñas, Y. (2005). Degradación y recuperación de los pastizales. Revista Cubana de Ciencia Agrícola, 39, 515–521.

Qin, X., Zhang, C., Sun, Z., Zhang, Y.-A., Song, R., & Du, M. (2018). Solving the multi-objective optimization model for alpine grassland grazing with modified genetic algorithms. In 2018 11th international congress on image and signal processing, biomedical engineering and informatics (cisp-bmei) (pp. 1–5).

Quaas, M., & Baumgärtner, S. (2012, 05). Optimal grazing management rules in semi-arid rangelands with uncertain rainfall. Natural Resource Modelling, 25. doi: 10.1111/j.1939-7445.2011.00110.x

Raizada, A., Dogra, P., & Dhyani, B. (2008). Assessment of a multi-objective decision support system generated land use plan on forest fodder dependency in a himalayan watershed. Environmental Modelling Software, 23(9), 1171-1181. Retrieved from https://www.sciencedirect.com/science/article/pii/S1364815208000182 doi: https://doi.org/10.1016/j.envsoft.2008.02.001

Romanzini, E. P., Watanabe, R. N., Fonseca, N. V. B., Berça, A. S., Brito, T. R., Bernardes, P. A., ... Reis, R. A. (2022). Modern livestock farming under tropical conditions using sensors in grazing systems. Scientific reports, 12(1), 1–10.

Tian, C., Manfei, X., Justin, T., Hongyue, W., & Xiaohui, N. (2018). Relationship between omnibus and post-hoc tests: An investigation of performance of the f test in anova. Shanghai archives of psychiatry, 30(1), 60.

White, R. R., Brady, M., Capper, J. L., & Johnson, K. A. (2014). Optimizing diet and pasture management to improve sustainability of u.s. beef production. Agricultural Systems, 130, 1-12. Retrieved from https://www.sciencedirect.com/science/article/pii/S0308521X14000766 doi: https://doi.org/10.1016/j.agsy.2014.06.004

Xiaoyan, G., Zirui, J., Yingying, Y., Dengsheng, W., & JianPing, L. (2019a). Research on dynamic rotational grazing assignment model based on grass resource leveling. Procedia Computer Science, 162, 696-703. Retrieved from https://www.sciencedirect.com/science/article/pii/S1877050919320502 (7th International Conference on Information Technology and Quantitative Management (ITQM 2019): Information technology and quantitative management based on Artificial Intelligence) doi: https://doi.org/10.1016/j.procs.2019.12.040

Xiaoyan, G., Zirui, J., Yingying, Y., Dengsheng, W., & JianPing, L. (2019b). Research on dynamic rotational grazing assignment model based on grass resource leveling. Procedia Computer Science, 162, 696–703.

Appendix F

An autonomous System for the self-supervision of animal fattening in the context of precision livestock farming

An Autonomous System for the Self-supervision of Animal Fattening in the Context of Precision Livestock Farming

Rodrigo García[1,2,6], Jose Aguilar[1,4,5], Angel Pinto[6]

[1]GIDITIC, Universidad EAFIT, Medellín, Colombia.

[2]Faculty of Engineering, Universidad de Córdoba, Montería, Colombia.

[4]CEMISID, Universidad de Los Andes, Merida, Venezuela.

[5]IMDEA Networks Institute, Leganés, Madrid, Spain.

[6]Faculty of Engineering, Universidad del Sinú, Montería, Colombia.

rjgarciah@eafit.edu.co, aguilar@ula.ve,angelpinto@unisinu.edu.co

Abstract

Beef production needs certain levels of autonomy to ensure that animal fattening processes achieve certain sustainability objectives (e.g., financial and environmental). For example, it is required oversight in the animal fattening process, so that stakeholders can make better decisions about what is happening in the fattening process. For monitoring the animal fattening process, this paper proposes an autonomous system. In this paper, this autonomous system is designed and implemented using the methodology for the development of data Mining applications called MIDANO, and is tested in a cattle farm simulator that has been developed to reproduce the events of the animal fattening production process. This autonomous system for the self-supervision of the animal fattening process is composed of two data analysis tasks, one to detect anomalies in the fattening of cattle, and another to diagnose this anomaly. The results with real data demonstrate the ability of the proposed supervision system to detect and diagnose anomalies in various conditions (normal, animal health problems, and forage problems in the paddock), and the possible causes of abnormal values in the weight variable. The anomaly detection models have a MAE of the order of 5.5 kilograms, and the diagnostic model has 95% of Accuracy and 1 of AUC. The results of the experiments are encouraging, as they show that the autonomous system is capable of detecting anomalies and diagnosing them in different operating scenarios. Our system allows giving self-supervision characteristics to a production process.

keywords: Autonomic Computing, Precision Livestock Farming, Artificial Intelligence, Machine Learning, Rotational Grazing, Beef Production

1 Introduction

To achieve sustainable development, natural resources must be treated with great care, and technology must be used to ensure the continued satisfaction of the human needs of present and future generations [1]. On the other hand, parallel to world population growth, the demand for animal protein is also increasing. Countries are reviewing the growth of animal production to meet the increasing demand for animal protein [2].

At the same time, autonomous computing is a paradigm with the main goal of defining computing systems that can manage themselves with little or no human interaction, with self-management capabilities to adapt to their environment [3]. In an autonomous computing environment, systems are able to monitor their own state, collect and analyze data, identify patterns and trends, and make intelligent decisions based on those analyses. They can adapt to changes in the environment, anticipate potential problems, and proactively take corrective action. The *autonomic cycle of data analysis tasks* (ACODAT) concept has been defined by Aguilar et al. [4–6] as a set of data analysis tasks that act autonomously to supervise and/or control a process. These Data Analysis tasks are based on knowledge models, and they interact and interrelate with each other according to the objectives of the cycle. Each data analysis task has a different function: to monitor the process, to analyze and interpret what is happening in the process, and to make decisions to improve the process.

On the other hand, Precision Livestock Farming (PLF) uses biometric sensors, artificial intelligence, big data and blockchain technologies. Biometric sensors include invasive or non-invasive sensors that monitor an individual animal's health and behavior in real-time, allowing producers to integrate this data for population-level analysis [7]. Particularly, PLF is a technological approach that uses various tools and techniques to monitor, control and optimize livestock production accurately and efficiently. To do this, it collects and processes a large amount of information about the animals, their environment and their performance, in order to make accurate decisions that improve production. The main objectives of PLF are [8]: to identify the most appropriate livestock feeding, to reduce environmental impact through efficient resource management, to manage crop processes to make a perfect synergy with livestock feeding, to ensure food safety through traceability of products, and to improve animal health and crop efficiency.

Also, beef production refers to the process of raising, feeding and managing cattle with the objective of obtaining quality meat for human consumption [9]. This process involves different stages, such as genetic selection of appropriate breeds, animal care in terms of feed, water, housing and veterinary care, as well as pasture management and proper nutrition.

In general, the livestock industry has numerous difficulties in detecting failures in its production systems, which directly affects ideal livestock growth and optimal paddock production. The problem lies in the inability to make decisions in real time, using data of what may be happening in the cattle fattening process. In a previous article, three ACODATs are proposed in the context of precision livestock farming to solve this problem [8].

This work defines and implements an ACODAT for the self-monitoring of animal fattening processes in the context of the PLF, with the objective of assisting farmers' decision making to increase production efficiency. The data analysis tasks in the ACODAT-based self-monitoring system play fundamental roles in the monitoring, analysis and decision making process in the rotational grazing system. The *first task* is responsible for collecting, cleaning and transforming the system and context data, ensuring that they are ready for further analysis (data preparation). The *second task* uses predictive models to detect anomalies in animal weights, comparing actual data with predictive model results and alerting to unexpected deviations (identification system). The *third task* uses classification techniques to determine the cause of the anomaly, which may be due to the health of the animal or the poor performance of the paddock (diagnostic system). Finally, the *last task* informs the farmer about the detected anomaly and its diagnosis, generating alarms or alerts that help him in decision making (notification). Specifically, the main contributions of this research are:

- The definition of the data analysis tasks for the self-supervision of the animal fattening process.

- The integration of various data mining techniques for the definition of knowledge models for beef production farms.

- The application of MIDANO for the definition of an ACODAT with the objective of monitoring and diagnosing the beef production process in order to enable real-time analysis and decision-making.

The scientific relevance of this article lies in its innovative proposal to apply the ACODAT concept in PLF, using advanced data mining techniques and MIDANO. Our approach allows real-time monitoring and diagnosis of the animal fattening process, improving the efficiency and productivity of beef production. The results and contributions of this work are relevant for the scientific community and professionals in the field of animal husbandry, offering an innovative tool for informed decision making and optimization of livestock production.

This article is organized as follows. Section 2 presents related work. Section 3 introduces the ACODAT for the supervision of cattle fattening proposed in this work. Section 4 shows the behavior of our approach through different case analyses in meat production. After, Section 5 describes the results obtained. Finally, Section 6 shows a comparison with previous works, and the article ends with conclusions in section 7.

2 Related Work

In the case of PLF, significant progress has been made in the use of tools to monitor animal health and production, which has generated a large amount of data [10]. Previous works in the field of PLF and beef production process monitoring have addressed different aspects related to livestock health and behavior. First, Sai *et al.* [11] and Haladjian *et al.* [12]. focused on detecting abnormalities in cattle health. Sai *et al.* [11] used environmental data and developed an artificial intelligence model to estimate the deep body temperature of animals, while Haladjian *et al.* [12] detected abnormalities in the movement pattern of cows as a possible indicator of lameness.

On the other hand, Wagner *et al.* [13] and Chung *et al.* [14] focused on activity monitoring and early detection of abnormalities. Wagner *et al.* [13] captured dairy cow activities and considered them as a time series for analysis, while Chung *et al.* [14] proposed a data mining solution for heat detection using sound data from native Korean cows. In addition, Kramer *et al.* [15] developed a fuzzy logic model for disease classification and monitoring in cattle, focusing on lameness and mastitis detection. On the other hand, Palomino et al. [16] focused on improving grazing systems for high mountain dairy cattle.

Finally, the papers [17,18] highlighted the importance of low-emission livestock farming in the context of climate change, proposing intensive grazing systems and considering the effects of climate change on the soil-plant-animal relationship. In addition, Garcia et al. [19] proposed an approach to detect anomalies in cattle fattening processes based on historical weight records.

Our work differs from previous works in several key aspects.

- First, while previous works focus on specific aspects of cattle monitoring, such as body temperature detection, movement pattern anomalies, or activity monitoring, our approach encompasses a completely autonomous system for monitoring the cattle fattening process based on rotational grazing.
- Furthermore, unlike previous works that focus on detecting anomalies in cattle health, our approach also considers other critical factors such as pasture conditions and environmental conditions. This allows us to have a comprehensive view of the fattening process, identifying possible deviations in both cattle health and environmental conditions.
- Another important difference is that we use data mining and fuzzy modeling techniques for the analysis and diagnosis of anomalies detected in the cattle fattening process. This allows us to gain a deeper understanding of the underlying causes of deviations, and provide more accurate recommendations for decision making.

The main contribution of our ACODAT is to provide an autonomous solution for monitoring the cattle fattening process based on rotational grazing. For that, it uses an identification model and a diagnostic system of the current situation to a better understanding of the fattening process. In addition, being an autonomous system, it allows a continuous and efficient supervision of the fattening process.

3 ACODAT-Based Supervision of Cattle Fattening

This section describes the proposed ACODAT-based self-supervision approach for rotational grazing beef cattle fattening systems. ACODAT is a novel and versatile concept that allows concurrent data-driven models to reach strategic goals. This concept has not been used previously in the context of animal fattening monitoring.

For the development of the architecture, the methodology MIDANO was used, which is used for the development of applications based on data analytics tasks [20, 21]. MIDANO is composed of three phases: identification of sources for knowledge extraction in an organization and definition of the ACODAT, data preparation and processing, and finally, implementation of the data analysis tasks.

3.1 ACODAT description

ACODAT is based on the autonomic computing paradigm [22], with the purpose of endowing autonomic properties to systems based on an intelligent control loop. The main goal of ACODAT is to extract useful knowledge from data to make decisions [4]. The set of data analysis tasks must be performed together to achieve the goal in the supervised process. The tasks are interconnected and have different roles in the loop, which are to observe the process, analyze and interpret what happens in it, and make decisions to achieve the objective for which the loop was designed. This closed loop approach allows the solution of complex problems.

ACODAT generally requires a multidimensional data model representing the data collected from the different sources, in order to characterize the behavior of the context. This model will be used for the different data analysis tasks.

3.2 Proposed general architecture

The rotational grazing system is composed of different tasks and decisions by the farmer, such as weighing the animals, defining what paddock to move them, checking the health status of the animals, determining the quality of the forage in the paddocks, among many others. The proposed autonomous cycle monitors the weighing of the animals, verifies the health status and the quality of the forage of the paddocks, using the data from the livestock software of the farm. The ACODAT-based self-supervision consists of four tasks, as shown in Figure 1.

In supervising the animal fattening process, Task 1 prepares the data, and Task 2 is a data analysis task that detects anomalies in animal weights. Task 3 is the other data analysis task that diagnoses the reason for this anomaly (animal health status or paddock performance), and Task 4 reports the anomaly and the diagnosis with possible causes for decision-making. Therefore, the functionalities provided by ACODAT-based self-supervision are as follows:

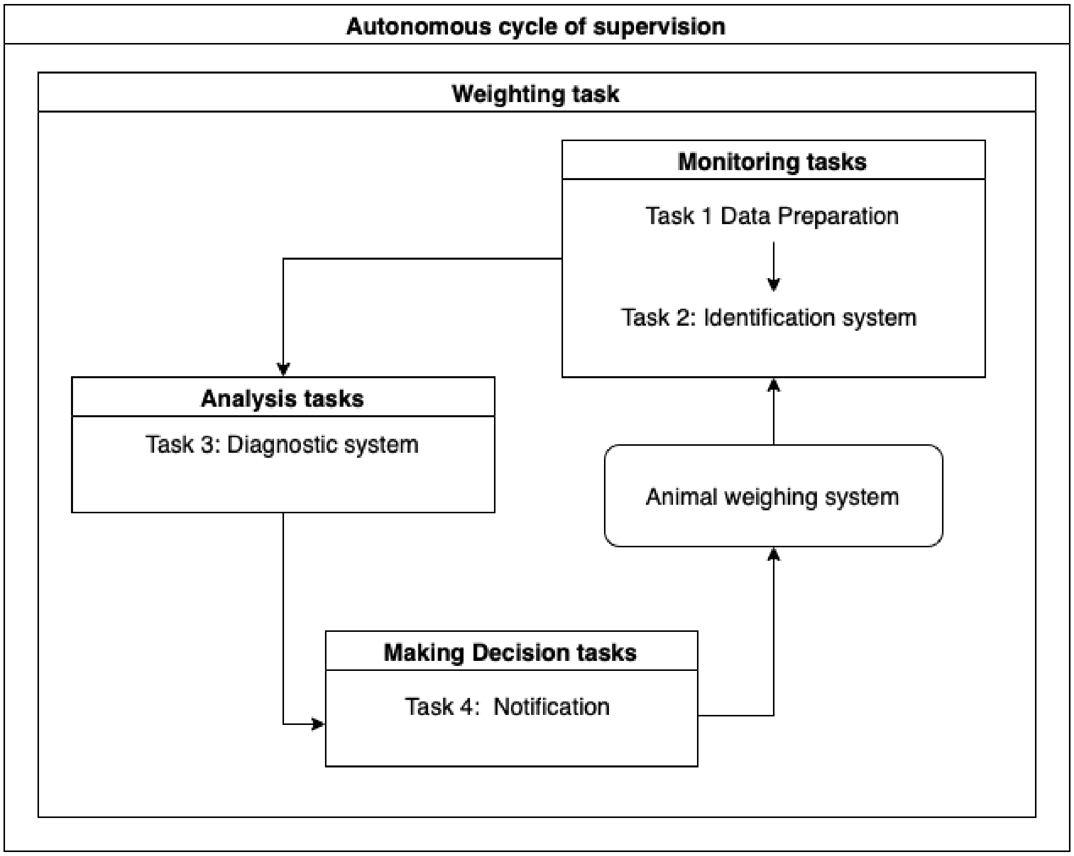


Figure 1: ACODAT-based supervision for PLF

- ***Monitoring process:*** a set of tasks (tasks 1 and 2) to capture data, and identify anomalies in the weight of the animals. In addition, they select and pre-process the relevant characteristics, and obtain information about animals behavior to be used in the steps below.

- ***Analysis process:*** is task 3 to interpret, understand and diagnose in real time what is happening in the fattening process.

- ***Decision-making process:*** this task defines and triggers notifications to the farmer. In addition, it could perform necessary physical actions on the controllable elements of the system. The effects of this task are returned for follow-up and analysis, restarting a new cycle.

Table 1 shows the proposed tasks, their roles in the ACODAT-based self-supervision system, and the data sources. The following subsections describe the set of tasks according to their role.

Table 1: ACODAT-based self-supervision tasks, functions and data sources for animal fattening process

Role	Task	Data Source
Monitoring	Task 1 Data Preparation	Animal Weighing System (Electronic Weighing Scale)
	Task 2: Identification system	Processed data from the previous task
Analysis	Task 3: Diagnostic system	Processed data from the previous task
Decision-Making	Task 4: Notification	Processed data from the previous task

3.3 Monitoring Role

This section describes the monitoring tasks of the ACODAT-based self-supervision system. According to Table 1, these tasks are the data preparation (Task 1) and the identification system (Task 2).

In particular, **Task 1** prepares the data, collecting them from the system and context, cleaning and transforming them, improving their quality. Also, this task may eventually consider virtual sensors, to predict some of the system/context variables. Therefore, it could eventually use predictive models. Table 2 describes the activities of Task 1. Specifically, some of the activities defined in this task are: Target variable selection, feature engineering, and data cleaning, among other data preparation processes.

Table 2: Description of Task 1 of data preparation.

Task 1	Data Preparation
Description:	Data collection
Data source:	Animal Weighing System
Data analytics type:	Maybe Prediction
Data analytics technique:	Correlation analysis Outlier detection
Knowledge model type:	Maybe Predictive model
Related data analytics task:	Identification system

Table 2: Description of Task 1 of data preparation.

Task 1	Data Preparation
Autonomic cycle type:	Monitoring

The monitoring process also provides anomaly detection. The objective of **Task 2** is a real-time analysis of the behavior of the weight variable, and detecting when it deviates from the ranges stipulated as normal. Thus, it immediately identifies the anomaly in the weight of the animals. The description of this task can be seen in Table 3.

Particularly, this task extracts knowledge for anomaly detection, for which it uses regression models. The regression model is not based entirely on the data but requires an expert to identify normal ranges of animal fattening. Once trained, the incoming data are compared with the results of the prediction model at a given time. Unexpected deviations between the two are an indication that an anomaly in animal weight is occurring.

Table 3: Description of Task 2 for the identification system

Task 2	Data Preparation
Description:	Anomaly detection
Data source:	Animal Weighing System
Data analytics type:	Prediction
Data analytics technique:	Decision tree Regression Random Forest
Knowledge model type:	Predictive model
Related data analytics task:	Diagnostic system
Autonomic cycle type:	Monitoring

There are different techniques to build the knowledge model for detecting anomalies, and it is necessary to select which to use for the case study, based on its performance in terms of prediction error. In previous work, this analysis has been carried out [19]. Overall, the results showed that the decision tree was the best performing, with a mean absolute error (MAE) of 5.4 kg.

3.4 Analysis Role

This section describes the task with this role in the ACODAT-based self-supervisory system to interpret and analyze the monitored information. According to Table 1, **Task 3** performs the diagnosis of animal fattening, that is, it determines the reason for the anomaly, as shown in Table 4. Its objective is to know if the loss of animal weight is due to the animal's health condition or to low yielding paddocks. In particular, this task defines a knowledge model to make a diagnosis of the anomaly.

Table 4: Description of Task 3 for the analysis.

Task 3	**Diagnostic system**
Description:	Diagnose the origin of the anomaly
Data source:	Previous task
Data analytics type:	Classification
Data analytics technique:	Fuzzy logic Genetic algorithms
Knowledge model type:	Diagnosis model
Related data analytics task:	Identification system
Autonomic cycle type:	Analysis

There are different techniques to build the knowledge model to diagnose anomalies, and it is necessary to select which one to use for the case study, based on its performance in terms of prediction error. In previous work, this analysis has been performed. In general, the results showed that the fuzzy classifier with rules optimized by genetic algorithms had good accuracy, when diagnosing weight loss by disease (97%), and in the case of diagnosing by paddock performance (95%) [23].

3.5 Decision-Making Role

In this section, we describe the task whose role in the ACODAT-based self-supervision system is to act from the anomaly detected and diagnosed in the previous phases. According to Table 1, **Task 4** is the notification of the current status of the animal fattening process. Specifically, Task 4 notifies the anomaly detected, with its diagnosis. It generates alarms or alerts caused by anomalous situations, such as a loss of weight, or a paddock with little forage, among other situations. The alarms are limited to warn about something, while the alerts not only warn but also request greater vigilance on something.

4 Case Study

For this case study, this section presents the experimental context and the instantiation of the ACODAT, using data from a beef farm located in Monteria-Colombia. The experimental context for this case study is as follows.

4.1 Experiment Context

In cattle ranching, a system widely used by beef producers in the tropics is rotational grazing. Rotational grazing consists of dividing the entire surface area of a farm into more than two paddocks, while some remain occupied, the others are at rest [24] (see Figure 2). This reduces the total grazing area, and forces cattle to consume forage uniformly, assigning batches of cattle to different paddocks [25]. Roughly speaking, this process consists of:

- Prepare the paddocks by sowing the pasture
- Buy the herds of animals
- Assign an initial paddock to each herd
- Each day x, assign a new paddock to each herd
- Regularly weigh the herds to measure their fattening.
- Sell the cattle when they reach a given weight

Some of the variables that can be considered in this process are the pasture as a source of feed for the cattle, the paddock capacity (amount of forage in each paddock), the animal's state of health, among others.

At a frequency stipulated by the farmer, each animal in the herd is weighed, in order to measure how much weight the animals have gained in the rotations carried out to measure and evaluate their production. At each herd weighing, the animal is identified and stored in the herd weighing software to have weight history. This is a repetitive process, since a farm may have many herds, and each herd may have different animals with different weights and breeds.

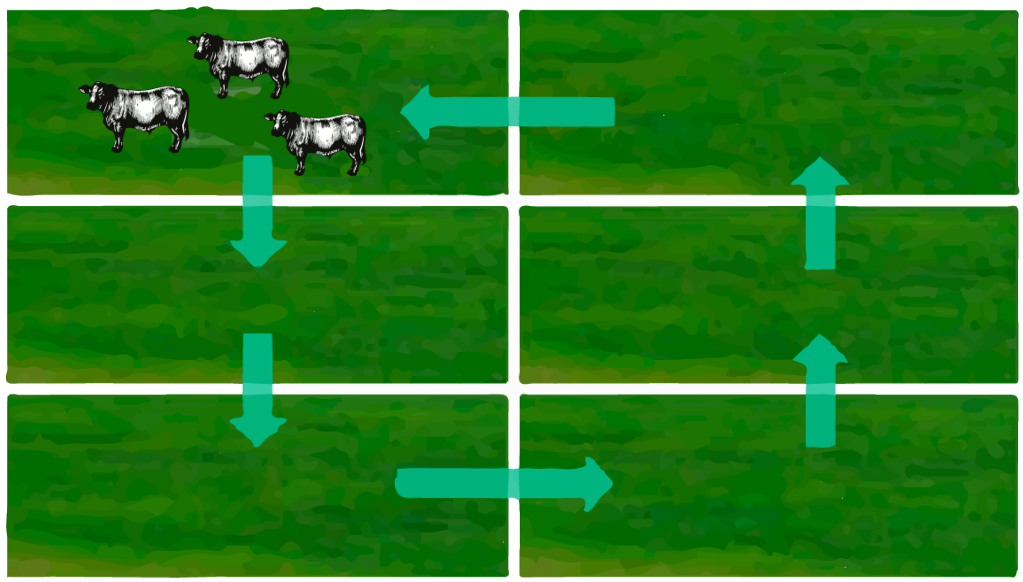


Figure 2: Rotational grazing system

4.2 Instantiation of ACODAT

This section describes the utilization of our ACODAT for the supervision of the animal fattening system at the Rosario farm. The Rosario farm has an average territorial extension with more than 100 paddocks, and works with various breeds of cattle, but the main ones are Angus, Angus Cebu, and Zebu. In general, the system starts with the weighing of the herds; immediately, the identification system detects an anomaly in the weight of an animal. If it is the case, it activates the diagnostic system, which will decide if this anomaly is due to the forage yield in the paddock or due to the state of animal health, as shown in Figure 3.

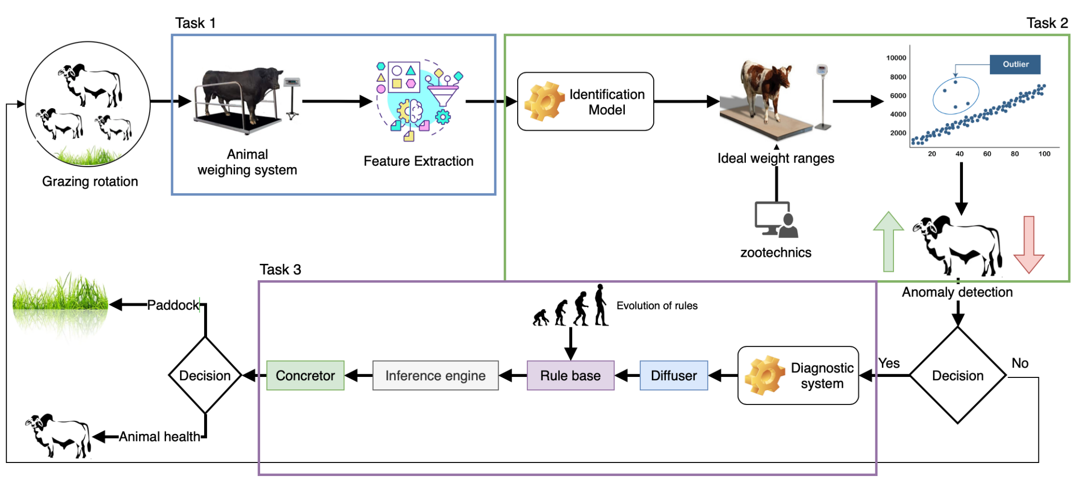


Figure 3: General system architecture based on ACODAT

The following subsections will explain the implementation of each proposed task in our architecture. The code and data used in the instances explained below, of the tasks described in Section 3, for the case study considered in this work, are found at https://github.com/devraxielh/PLF-Identification-Diagnostic.

4.2.1 Task 1: Preparation of the Data.

Data extraction is straightforward and only requires collection and understanding. The compilation is carried out on a database made up of several tables that are the result of pre-existing queries on some variables chosen with different exploration frequencies. The most significant table is the history of the weights, which come directly from the electronic scale. On the other hand, feature engineering was used where for the extraction and description of the relevant features; in this case, age, race, and gender, were identified.

4.2.2 Task 2: Anomaly detection

This task is defined by an identification model that can be used in different contexts (e.g., control or monitoring process). The objective is to develop a predictive model that can identify unusual changes in animal weight during the fattening process. The identification process determines the ideal range of weight values for cattle. With this range, weights considered normal or atypical are established, using a forest isolation algorithm that returns an anomaly score in relation to the expected weight of the animal. In this way, anomalous shifts in weight up or down can be detected. The model is used for anomaly detection in the beef production process. Thus, the identification model was defined as a predictive model, and was built using machine-learning techniques from data collected and validated by animal technicians at the Rosario farm.

The identification process determines the ideal range of values of cattle weights. With this range, the expected or considered normal or atypical weights are determined, by implementing a forestall isolation algorithm that returns the score of the anomaly with respect to the expected weight of the animal, thus detecting anomalous weight changes up or down.

We compare different prediction techniques such as Decision Tree (DT), Gradient boosting (GB), K-neighbors(KN) and Random Forest (RF), in order to model the ideal weight of an animal as a function of time, breed and gender. These algorithms are used because they are efficient in regression tasks and have shown to be effective in solving a wide variety of problems [26–28]. For each model, a hyperparameter optimization process was carried out, to choose in each case the optimal set of hyperparameters for each learning algorithm. In the case of the approaches based on decision trees (DT, GB and RF) the depth of the tree (4-10) is optimized, and the function to divide the data (mean square error, etc.); in the case of KN, the value of K (3 to 10) and the similarity function used (Euclidean distance, Minkowski distance, Manhattan distance, etc.). The dataset contains the records of 104 animals from the day of birth to slaughter, and it is used a ten fold cross-validation strategy (80% of the records for training and 20% for testing). Figure 4 shows the mean absolute error (MAE) of weight prediction for each technique used, which is calculated using equation 1.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \widehat{y}_i| \tag{1}$$

where N is the total number of records in the dataset, y_i and $\widehat{y}_i$ are the current weight and predicted weight by the model, respectively.

According to Figure 4, DT and RF show very similar behaviors, and the lowest MAEs with medians below 5.5 kilograms. In general, in Table 5, the mean MAE of DT and RF are very similar and present the smallest values, thus confirming that they are the techniques with the lowest prediction error, but there are no significant differences between them. However, the mean training time of DT is 0.0651 seconds, while that for RF is 0.445 seconds. Therefore, DT is as suitable as RF for the identification of the ideal weight of the animals, but it is much faster (almost 7 times).

Table 5: Comparison of results

Technique	**DT**	**RF**	**GB**	**KN**
Mean MAE	5.4525	5.4536	5.7249	5.9714

In this context, an MAE of 5.5 kilograms indicates that, on average, the predictions from our anomaly detection model may differ from the actual values by about 5.5 kilograms. The lower the MAE value, the higher the accuracy of the model, since it implies that the predictions are closer to the actual values. Therefore, in this specific case, an MAE of the order of 5.5 kilograms indicates that the anomaly detection models have an acceptable accuracy in estimating deviations in the weight of the animals (which fluctuate by about 350 kilograms on average during the fattening process), which allows them to identify and alert on possible anomalies in the fattening process.

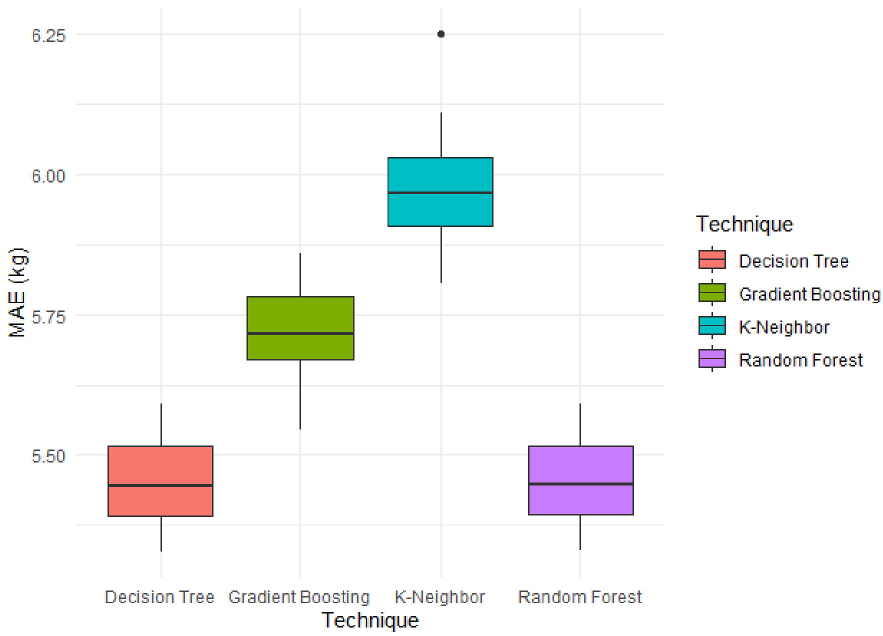


Figure 4: MAE boxplot for each technique.

4.2.3 Task 3: Diagnostic system

In the context of diagnosing the animal fattening process in a grazing system, fuzzy logic can be especially useful due to the inherently imprecise and subjective nature of certain aspects of the system. For example, forage quality and forage tolerance to weather conditions can be evaluated in terms of grades, such as "low", "medium" or "high". Fuzzy logic allows representing and working with these degrees in a more adequate and flexible way than classical logic, as has been demonstrated in previous successful works in different fields [29–32].

This task aims to diagnose a detected anomaly. The diagnostic system is based on a fuzzy classifier that uses fuzzy rules to characterize the diagnostic process, and fuzzy reasoning to determine the current situation before a given input. In addition, the fuzzy classifier optimizes the rules using genetic algorithms, which modify the membership functions, providing a more accurate system for diagnosis.

The rules are designed and created using a fuzzy clustering algorithm. The fuzzy-c-means (FCM) algorithm is used, which allows fuzzy data clustering (each cluster describes a fuzzy set for the linguistic/fuzzy variable of interest). In addition, they are optimized using a Genetic Algorithm (GA), whose task is to improve the definition of the fuzzy sets (it adapts the membership function of each fuzzy set).

The first step in designing the fuzzy system consists of determining the fuzzy/linguistic variables and generating their fuzzy sets. The FCM algorithm allows defining the fuzzy sets from the data of each linguistic variable. The linguistic variables used are those linked to the fattening process. Particularly, the input variables are the age, the paddock, and the differential of the final weight with the initial weight, and the output variable is the state of the animal.

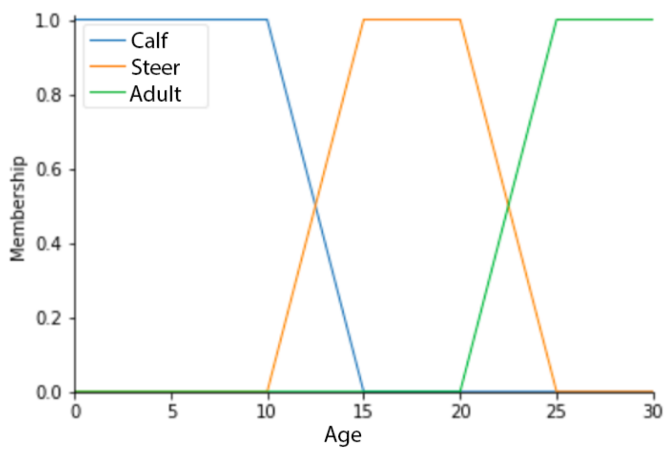


Figure 5: Fuzzy Variable of Age

Age: This variable is shown in Figure 5, its universe of speech is $[0, 30]$, and represents months of life. Its fuzzy sets are $[calf, steer, adult]$.

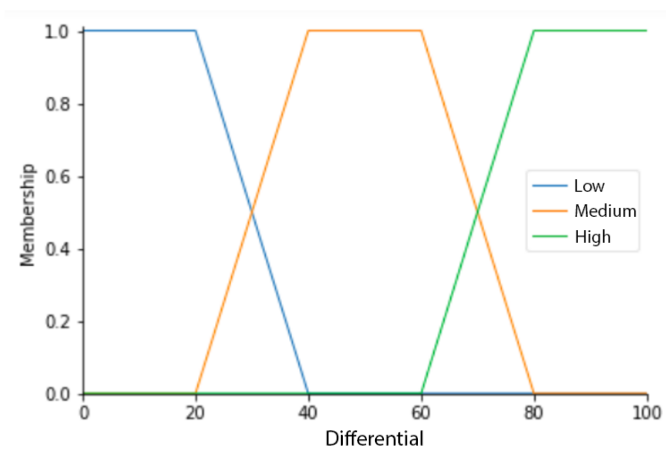


Figure 6: Fuzzy Variable of Differential

Differential: This variable is shown in Figure 6, its universe of discourse is $[0, 100]$, and represents the percentage of the difference between the initial and current weight. Its fuzzy sets are $[low, medium, high]$. The differential variable allows measuring and monitoring the impact of animal rotation on animal weight. If the weight difference is positive, then it means that the animals have gained weight during the grazing cycle, indicating good performance. On the other hand, if the weight difference is negative, then it means that the animals have lost weight, which may indicate health problems or poor pasture performance in the paddock.

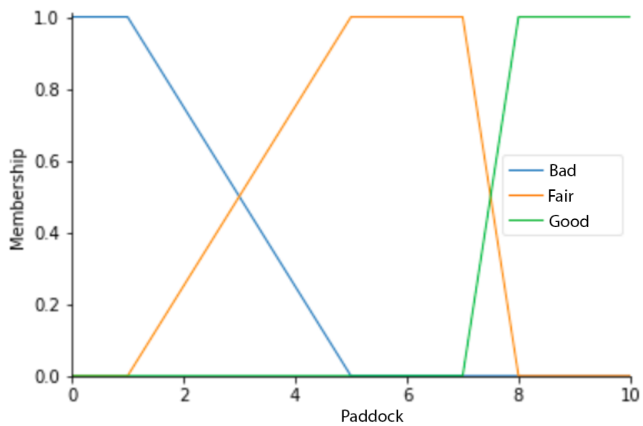


Figure 7: Fuzzy Variable of Paddock

Paddock: this variable is shown in Figure 7, its universe of discourse is $[0, 10]$, and represents the paddock yield score, characterized by forage quality and forage tolerance to the climatic situation. Its fuzzy sets are $[bad, fair, good]$. The paddock variable allows evaluating and selecting the most suitable paddocks to move the animals based on their performance. A paddock with a high performance score would indicate that it has high quality forage and a higher tolerance to weather conditions, which would be beneficial for fattening the animals. On the other hand, a paddock with a low yield score could indicate that the forage is of low quality or that the paddock is not in optimal condition due to adverse climatic factors.

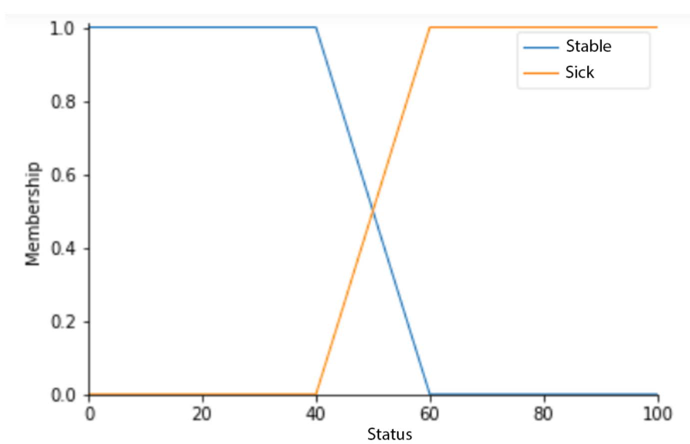


Figure 8: Fuzzy Variable of Status

Animal Status: this variable is shown in Figure 8, its universe of discourse is $[0, 100]$, represented by a percentage to indicate how well the animal is. Its fuzzy sets are $[stable, sick]$.

The rules that model the system are of the form $IF < ... > THEN < ... >$. In the following tables, we define the set of rules.

Table 6: Rules (fuzzy system) with 2 antecedents

		Differential		
		low	medium	high
Age	Calf	sick	sick/stable	stable
	Steer	sick	sick/stable	stable
	Adult	sick/stable	sick/stable	stable

Table 7: Rules (fuzzy system) with 3 antecedents

		Differential				
		low	medium	high		
Age	Calf	sick	sick/stable	stable	fair/bad	paddock
	Steer	sick	sick/stable	stable		
	Adult	sick/stable	sick/stable	stable		

Thus, there are two different types of fuzzy rules, depending on whether information on paddock performance is available. We have a group of rules with two antecedents (see Table 6) or with three antecedents (see Table 7). Particularly, in Table 6, the two antecedents are the differential and age variables; and in Table 7, the same but now also the paddock variable (the values of it that directly influence the animal fattening, such as regular/bad). An example of a rule for the latter case is:

If the age is calf and the differential is low and the paddock is bad, then the animal's condition is sick

These rules will be fitted/optimized to the input dataset using Genetic Algorithms. Specifically, the membership functions of the fuzzy sets will be the ones fitted according to the input data. A fitness function based on the calculation of the following measures was used:

- *Accuracy*: measures the accuracy of an individual. In this case, as it is based on labeled data on diagnosis, the score establishes whether the consequent is true (rule is triggered) when it should. It is calculated by the following equation:

$$Accuracy = \frac{tp}{(tp + tf)} \tag{2}$$

  Where pv is the number of true positives, and pf is the number of false positives. In particular, the score for each rule is calculated for the input dataset, and averaged.

- *Trigger degree of each rule (certainty):* it is the activation level of a rule, calculated as the product of the membership functions of the antecedents. If the value is high, then it means that the activation level of the rule is high.

At the same time, an individual is defined as a set of possible membership functions for each fuzzy variable. Finally, the fitness function of each individual is the average score of each rule multiplied by its degree of certainty. Thus, the evolutionary process, using this hybrid fitness function (score and degree of certainty), allows adjusting the fuzzy sets to adapt them to the dataset.

To adjust the fuzzy system, two case studies were considered. The first case study is where at least one animal is labeled sick, and paddock performance is good despite the weather season. The rules that will be activated are those of two antecedents since the paddock performance is good and does not negatively affect animal fattening.

The second case study presents a variation, which consists of analyzing animal fattening considering the performance of the paddock when it is regular. The rules that will be activated are those of 3 antecedents, in this case, the paddock variable is used because it does affect negatively animal fattening.

The diagnostic system is evolved and the results with the best individual (quality metrics) are shown in Tables 8 and 9. We have added the AUC metric, which stands for "Area under the ROC (receiver operating characteristic) Curve". An ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds.

An AUC value equal to 1 is excellent, as is accuracy equal to 100 or a rule certainty equal to 1. The results of the best individual in each case are very good because these values are close to or equal to the previous values.

Table 8: Case 1 results

Metrics	**Values**
Accuracy	97%
Certainty of rules	R1:0.85, R2:0.88,R3:0.90,R4:0.85,R5:0.85
AUC	1

Table 9: Case 2 results

Metrics	**Values**
Accuracy	95%
Certainty of rules	R1:0.88, R2:0.75, R3:0.89, R4:0.85, R5:0.85, R6:0.75
AUC	1

5 Experiments

In our ACODAT, there are two main data analysis tasks: one for the definition of the cattle fattening process identification model, and another for the diagnosis of anomalies in a beef production process. We will use 3 test scenarios: normal, animal health problem, and forage problem in the paddock. In this section, we are going to describe the behavior of ACODAT previously trained, using real data from the "El Rosario Cattle" farm located in Monteria-Colombia.

5.1 A normal scenario

Input: A herd of 15 animals with an average age of 2 years of zebu breed.

ACODAT analysis: The 15 animals are weighed on the farm scale (animal weighing system). Immediately, task 1 is activated to prepare the data, then these data are passed to task 2 where the anomaly detection is carried out, and will return the status of the herd (see Figure 9).

Output: In this case, Task 3 would not be carried out because the herd of animals gained the desired weight. Therefore, they return to rotational grazing.

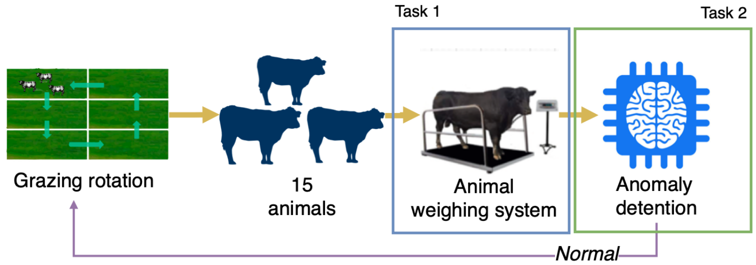


Figure 9: Normal scenario

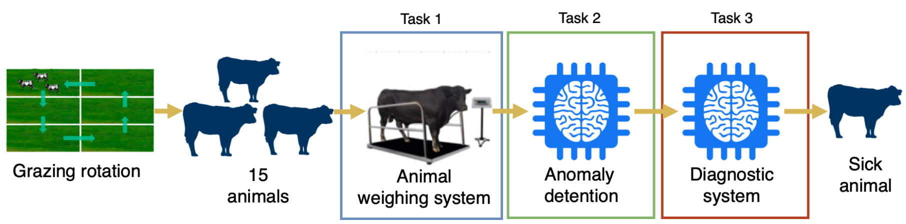


Figure 10: Sick animal scenario

5.2 Sick animal scenario

Input: A herd of 7 animals with an average age of 1.6 years of AngusxCebu (AC) breed.

ACODAT analysis: Task 2 detects that one animal in a herd of 7 animals with an average age of 1.6 years did not gain the appropriate weight for its age (see Figure 10). Particularly, it compares the ideal weight growth curve described by the identification model (see Figure 11) with the animal weights (point cloud). Since the curve represents the ideal temporal behavior of the weight, points that deviate from it are potentially anomalous, but it is necessary to define the magnitude of the distance that separates an anomalous weight from one considered normal. To do this, multiple predictions of ideal weights are obtained for each combination of day, race, and sex. Next, the ideal (non-anomalous) weight intervals are constructed to detect anomalous weights [19] (see Table 10).

Table 10: Detection of anomalous data in a sick animal scenario

Days	Actual Weight	Breed	Gender	Ideal Weights	State
600	345	AC	M	[328-380]	normal
600	350	AC	M	[328-380]	normal
600	360	AC	M	[328-380]	normal
600	337	AC	M	[328-380]	normal
600	315	AC	M	[328-380]	abnormal

600	345	AC	M	[328-380]	normal
600	332	AC	M	[328-380]	normal

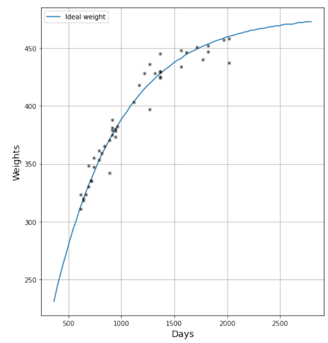


Figure 11: Ideal weight-curve versus real weight [19]

Subsequently, Task 3 is invoked with existing fuzzy rules. The fuzzy classifier system infers that only one animal is sick for that given input.

The first fuzzy rules are used (see Tables 6-7). Then, once several iterations of the fuzzy system have been performed, the adaptive system of rules is invoked, which is based on a Genetic Algorithm. An example of this evolution is shown. For example, in Figure 12, the rules of the 1st generation do not satisfy the system since they are rules with stable animal state. In the 2nd iteration, an improvement in the rules is noticed because the sick state appears in the consequent. Finally, in the 4th iteration, the stop condition is reached, and its best individual (set of rules) represents the final solution (new diagnosis system).

1	2
Rules: ...... IF age[high] AND dif[high] THEN status[stable] IF age[steer] AND dif[high] THEN status[stable]	Rules: ...... IF age[steer] AND dif[low] THEN status[sick] IF age[adult] AND dif[high] THEN status[stable] IF age[steer] AND dif[high] THEN status[stable]
3	**4**
Rules: ...... IF age[steer] AND dif[low] THEN status[sick] IF age[adult] AND dif[high] THEN status[stable] IF age[steer] AND dif[high] THEN status[stable] IF age[steer] AND dif[medium] THEN status[sick]	Rules: ...... IF age[steer] AND dif[low] THEN status[sick] IF age[adult] AND dif[high] THEN status[stable] IF age[steer] AND dif[high] THEN status[stable] IF age[calf] AND dif[high] THEN status[stable] IF age[calf] AND dif[medium] THEN status[sick]

Figure 12: Evolution of the rules that are activated

Output: The system diagnose that the problem is due to animal health since it was the only animal in the herd that did not gain the weight considered normal for its age. Therefore, a decision is made regarding the animal's health so that it can return to rotational grazing.

5.3 Scenario of a paddock with low yield

Input: A herd of 10 animals with an average age of 1.6 years of Angus (AN) breed.

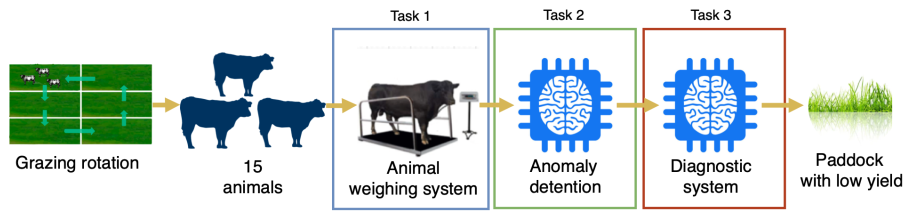


Figure 13: Scenario of a paddock with a low yield

ACODAT analysis: Task 2 with the anomaly detection model detects that 7 animals from a herd with 10 animals with an average age of 1.6 years did not gain the appropriate weight for their age (see Table 11).

Table 11: Detection of anomalous data in a scenario of a paddock with low yield

Days	Actual Weight	Breed	Gender	Ideal Weights	State
600	328	AN	M	[332-387]	abnormal
605	330	AN	M	[336-391]	abnormal
600	310	AN	M	[332-387]	abnormal
600	334	AN	M	[332-387]	normal
600	333	AN	M	[332-387]	normal
600	337	AN	M	[332-387]	normal
600	321	AN	M	[332-387]	abnormal
605	327	AN	M	[336-391]	abnormal
600	329	AN	M	[332-387]	abnormal
603	328	AN	M	[333-389]	abnormal

It makes the diagnostic request to Task 3, which is invoked with the existing fuzzy rules. The fuzzy classifier system infers that several animals are sick for that given input, and that the problem is with paddock performance. Again, once several iterations of the fuzzy system have been performed, the adaptive rule system is invoked.

Output: It diagnoses that the problem is with paddock performance since on average, most of the animals in the herd did not gain the weight considered normal for their age. Therefore, a decision is made regarding paddocks.

6 Comparison with other works

In this section, we propose criteria to compare our approach with similar works. In this section, we have focused on determining whether the works comply with some of these aspects, and not on how they do so methodologically. To compare this work with other similar works, we proceeded to define four criteria, which are:

- ***Criterion 1:*** The work jointly analyzes grazing and animal welfare.
- ***Criterion 2:*** The beef production process is supervised by an autonomous system.
- ***Criterion 3:*** ML techniques are used to detect anomalies in animals.
- ***Criterion 4:*** ML techniques are used to diagnose anomalies in a beef production process.

In Table 12, a qualitative comparison with related studies is made, based on previous criteria.

Table 12: Comparison with previous works.

	Criterion 1	***Criterion 2***	***Criterion 3***	***Criterion 4***
[11]	✗	✗	✓	✗
[12]	✗	✗	✓	✗
[13]	✗	✗	✗	✓
[14]	✗	✗	✓	✗
[15]	✗	✗	✗	✓
[16]	✓	✗	✗	✗
[17]	✓	✗	✗	✗
[18]	✓	✗	✗	✗
[19]	✗	✗	✓	✗
This work	✓	✓	✓	✓

As shown in Table 12, the related works do not meet all criteria. Specifically, in criterion 1, our work considers the diagnosis of animal health status, which contributes to animal welfare, and diagnosis of paddocks, which contributes to pasture improvement. In both cases, the farmer can intervene to improve these conditions. Palomino et al. [16] determined the ideal number of paddocks and the optimal stocking rate. In addition, [17, 18] consider the effects of climate change on the soil-plant-animal relationship and propose an intensive grazing system that guarantees low-emission livestock farming.

Regarding criterion 2, none of the related works use autonomous systems for fattening monitoring in livestock production. Our proposal is based on ACODAT principles. This research allows, through ACODAT, self-monitoring of the detection and diagnosis of anomalies in the meat production process by rotational grazing.

In relation to criterion 3, Sai et al. [11] performed the detection of anomalies in cattle health. Haladjian et al. [12] detected abnormalities in a cow's movement pattern as a possible indicator of lameness. Chung et al. [14] proposed a data mining solution for heat detection using sound data from native Korean cows. Finally, Garcia et al. [19] proposed an approach to detect anomalies in the cattle fattening process using historical records of animal weights. Our approach uses anomaly detection as input to the diagnostic model, identifying whether animals have gained the appropriate weight over time.

With respect to criterion 4, Wagner et al. [13] diagnosed the status of cows (disease, stress, no problems) using a ruminal sensor. In addition, Kramer et al. [15] developed a fuzzy logic model for classification and monitoring of lameness and mastitis in cows using data. Finally, our proposal focuses on the diagnosis of the detected anomaly, with the aim of achieving intelligent self-monitoring of animal fattening in rotational grazing.

Thus, the only work that meets all four criteria is ours, which is suitable for PLF environments, allowing self-adaptive, reconfigurable, scalable, high-precision systems. Also, an interesting aspect is how to extend them to multi-agent approaches, for natural modeling of the problem [33]. In general, our approach can easily accommodate a multi-agent environment, defining animals, grass, etc., as agents of the system.

7 Conclusions

This paper proposes a novel self-supervision system for the detection and diagnosis of anomalies present in animal weights in a fattening system based on rotational grazing. The work is based on the ACODAT concept and applies it to PLF. The ACODAT concept has been successfully tested in different areas [34–36], but has not yet been used in PLF [37]. Our ACODAT-based self-supervision system is capable of detecting anomalies and carrying out the diagnosis of what is happening on a beef farm. The ACODAT tasks are based on several ML techniques working together with specific objectives such as anomaly detection and diagnosis in rotational grazing.

Our ACODAT is composed of 4 tasks that allow real-time monitoring of the animal fattening process, detecting and diagnosing anomalies, and notifying the farmer so that he can make informed decisions. The data preparation task ensures the quality and relevance of the information used in the analysis, while the identification and diagnostic tasks use modeling techniques to identify and understand anomalies. Finally, the reporting task provides clear and timely communication to the farmer, providing crucial information for proper decision making in the management of the rotational grazing system. Together, these tasks strengthen the autonomous monitoring process and improve efficiency and productivity in farming practice.

The proposed autonomous cycle was tested with real data from the "El Rosario Cattle" farm located in Monteria-Colombia. The rotational grazing system is one of the most used in the tropics for cattle fattening [9], which makes the scenario very appropriate to extend the results to other farms. The results with real data demonstrate the ability of the proposed supervision system to detect and diagnose anomalies in various conditions (normal, animal health problems, and forage problems in the paddock), and the possible causes of abnormal values in the weight variable. This scheme using ACODAT provides a novel approach to anomaly detection and diagnosis in PLF. In addition to the flexibility to select techniques and model configurations, the ACODAT supervisory scheme can be retrained in real-time, increasingly adapting to the supervised system and improving its prediction accuracy.

Future work will extend this ACODAT-based self-supervision system to other types of production such as dairy cows. Other future work is to extend the current data with information on more anomalies in order to expand the capability of our system to diagnose a larger number of anomalies. In addition, other future work will incorporate meta-learning approaches to autonomously update cycle knowledge models, or to transfer them to other farms, among other things. Finally, future work should carry out further testing of our approach in different geographic locations and climatic conditions, to show the extensibility of the approach, preliminary results show that this idea should work.

Declarations

Data Availability Statement

Data and code of this work is available in:

https://github.com/devraxielh/PLF-Identification-Diagnostic.

Funding and/or Conflicts of interests/Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Rodrigo García is supported by a PhD grant financed by Universidad EAFIT. All the authors would like to thank the "Vicerrectoría de Descubrimiento y Creación", of Universidad EAFIT, for their support on this research. Also, the authors want to thank the HAMADI 4.0 project, code 22-STIC-06, of the STIC-AmSud regional program, for the support of this work. Finally, the authors want to thank the student Manuela Zapata Mesa, from the Department of Mathematical Sciences, EAFIT University, for the scientific assistance provided during the realization of this work.

References

[1] L. O. Tedeschi, J. P. Muir, D. G. Riley, and D. G. Fox, "The role of ruminant animals in sustainable livestock intensification programs," International Journal of Sustainable Development & World Ecology, vol. 22, no. 5, pp. 452–465, 2015.

[2] H. Erdal, G. Erdal, and B. Ayyildiz, "Are support policies for sustainable livestock important? causality between animal existence and support policies: Vecm analysis for turkey," Journal of Animal and Plant Sciences, vol. 31, no. 1, pp. 254–264, 2021.

[3] P. J. Horn, "Autonomic computing: Ibm's perspective on the state of information technology," 2001.

[4] J. Aguilar, O. Buendia, A. Pinto, and J. Gutiérrez, "Social learning analytics for determining learning styles in a smart classroom," Interactive Learning Environments, vol. 30, no. 2, pp. 245–261, 2022.

[5] M. Sánchez, J. Aguilar, J. Cordero, P. Valdiviezo-Díaz, L. Barba-Guamán, and L. Chamba-Eras, “Cloud computing in smart educational environments: Application in learning analytics as service,” in New Advances in Information Systems and Technologies (Á. Rocha, A. M. Correia, H. Adeli, L. P. Reis, and M. Mendonça Teixeira, eds.), pp. 993–1002, 2016.

[6] J. Vizcarrondo, J. Aguilar, E. Exposito, and A. Subias, “Mape-k as a service-oriented architecture,” IEEE Latin America Transactions, vol. 15, no. 6, pp. 1163–1175, 2017.

[7] E. Tullo, I. Fontana, A. Diana, T. Norton, D. Berckmans, and M. Guarino, “Application note: Labelling, a methodology to develop reliable algorithm in plf,” Computers and Electronics in Agriculture, vol. 142, pp. 424–428, 2017.

[8] R. García, J. Aguilar, M. Toro, N. Pérez, A. Pinto, and P. Rodríguez, “Autonomic computing in a beef-production process for precision livestock farming,” Journal of Industrial Information Integration, vol. 31, p. 100425, 2023.

[9] O. Feitosa, R. Leite, E. Alexandrino, T. Pires, L. de Oliveira, J. Neto, and A. Dos Santos, “Forage performance and cattle production as a function of the seasonality of a brazilian tropical region,” Acta Scientiarum - Animal Sciences, vol. 44, 2022. cited By 0.

[10] G. Morota, R. V. Ventura, F. F. Silva, M. Koyama, and S. C. Fernando, “Big data analytics and precision animal agriculture symposium: Machine learning and data mining advance predictive big data analysis in precision animal agriculture1,” Journal of Animal Science, vol. 96, pp. 1540–1550, 03 2018.

[11] S. Ma, Q. Yao, T. Masuda, S. Higaki, K. Yoshioka, S. Arai, S. Takamatsu, and T. Itoh, “Development of an anomaly detection system for cattle using infrared image and machine learning,” Sens. Mater., vol. 32, no. 12, pp. 4139–4149, 2020.

[12] J. Haladjian, Z. Hodaie, S. Nüske, and B. Brügge, “Gait anomaly detection in dairy cattle,” in Proceedings of the Fourth International Conference on Animal-Computer Interaction, ACI2017, (New York, NY, USA), Association for Computing Machinery, 2017.

[13] N. Wagner, V. Antoine, J. Koko, M.-M. Mialon, R. Lardy, and I. Veissier, “Comparison of machine learning methods to detect anomalies in the activity of dairy cows,” in Foundations of Intelligent Systems (D. Helic, G. Leitner, M. Stettinger, A. Felfernig, and Z. W. Raś, eds.), (Cham), pp. 342–351, Springer International Publishing, 2020.

[14] Y. Chung, J. Lee, S. Oh, D. Park, H. Chang, and S. Kim, "Automatic detection of cow's oestrus in audio surveillance system," Asian-Australasian journal of animal sciences, vol. 26, no. 7, p. 1030–1037, 2013.

[15] E. Kramer, D. Cavero, E. Stamer, and J. Krieter, "Mastitis and lameness detection in dairy cows by application of fuzzy logic," Livestock Science, vol. 125, no. 1, pp. 92–96, 2009. cited By 49.

[16] F. E. Palominos, H. M. Díaz, S. K. Palominos, and L. R. Cañete, "Relación entre los procedimientos de selección a la educación superior y el desempeño académico de los estudiantes con base en una clasificación mediante conjuntos difusos," Formación universitaria, vol. 11, no. 1, pp. 45–52, 2018.

[17] J. O. Chelotti, Sistema de adquisición y análisis de información acústica para ganadería de precisión. PhD thesis, 2018.

[18] Y. Qiao, M. Truman, and S. Sukkarieh, "Cattle segmentation and contour extraction based on mask r-cnn for precision livestock farming," Computers and Electronics in Agriculture, vol. 165, p. 104958, 2019.

[19] R. García, J. Aguilar, M. Toro, and M. Jiménez, "Weight-identification model of cattle using machine-learning techniques for anomaly detection," in 2021 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 01–07, IEEE, 2021.

[20] J. Vizcarrondo, J. Aguilar, E. Exposito, and A. Subias, "Armiscom: Autonomic reflective middleware for management service composition," in 2012 Global Information Infrastructure and Networking Symposium (GIIS), pp. 1–8, 2012.

[21] F. Pacheco, C. Rangel, J. Aguilar, M. Cerrada, and J. Altamiranda, "Methodological framework for data processing based on the data science paradigm," in 2014 XL Latin American Computing Conference (CLEI), pp. 1–12, 2014.

[22] J. O. Kephart and D. M. Chess, "The vision of autonomic computing," Computer, vol. 36, no. 1, pp. 41–50, 2003.

[23] C. Benitez, R. García, J. Aguilar, M. Jiménez, and H. Robles, "Supervision system of the fattening process of cattle in rotational grazing using fuzzy classification," in 2022 XVLIII Latin American Computer Conference (CLEI), pp. 1–10, 2022.

[24] D. Briske, J. Derner, J. Brown, S. Fuhlendorf, W. Teague, K. Havstad, R. Gillen, A. Ash, and W. Willms, “Rotational grazing on rangelands: Reconciliation of perception and experimental evidence,” Rangeland Ecology Management, vol. 61, no. 1, pp. 3–17, 2008.

[25] E. J. Jacobo, A. M. Rodríguez, N. Bartoloni, and V. A. Deregibus, “Rotational grazing effects on rangeland vegetation at a farm scale,” Rangeland Ecology Management, vol. 59, no. 3, pp. 249–257, 2006.

[26] J. R. Quinlan, “Decision trees and decision-making,” IEEE Transactions on Systems, Man, and Cybernetics, vol. 20, no. 2, pp. 339–346, 1990.

[27] M.-J. Jun, “A comparison of a gradient boosting decision tree, random forests, and artificial neural networks to model urban land use changes: The case of the seoul metropolitan area,” International Journal of Geographical Information Science, vol. 35, no. 11, pp. 2149–2167, 2021.

[28] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, “A comparative analysis of gradient boosting algorithms,” Artificial Intelligence Review, vol. 54, pp. 1937–1967, 2021.

[29] O. Abu Arqub, “Adaptation of reproducing kernel algorithm for solving fuzzy fredholm–volterra integrodifferential equations,” Neural Computing and Applications, vol. 28, pp. 1591–1610, Jul 2017.

[30] O. Abu Arqub, S. Momani, S. Al-Mezel, and M. Kutbi, “Existence, uniqueness, and characterization theorems for nonlinear fuzzy integrodifferential equations of volterra type,” Mathematical Problems in Engineering, vol. 2015, p. 835891, Mar 2015.

[31] O. Abu Arqub, J. Singh, B. Maayah, and M. Alhodaly, “Reproducing kernel approach for numerical solutions of fuzzy fractional initial value problems under the mittag–leffler kernel differential operator,” Mathematical Methods in the Applied Sciences, vol. 46, no. 7, pp. 7965–7986, 2023.

[32] O. Abu Arqub, J. Singh, and M. Alhodaly, “Adaptation of kernel functions-based approach with atangana–baleanu–caputo distributed order derivative for solutions of fuzzy fractional volterra and fredholm integrodifferential equations,” Mathematical Methods in the Applied Sciences, vol. 46, no. 7, pp. 7807–7834, 2023.

Appendix G

Management System for the Fattening Process of Bovines in Rotational Grazing using Diagnosis and Recommendation Systems

Management System for the Fattening Process of Bovines in Rotational Grazing using Diagnosis and Recommendation Systems

Rodrigo García
GIDITIC, Universidad EAFIT
Dpto Systems, Universidad del Córdoba
Dpto Systems, Universidad del Sinú
Monteria, Colombia
0000-0002-5935-8862

Charles Benitez
Faculty of Engineering
Universidad de los Andes
Merida, Venezuela
0000-0001-9975-4790

Jose Aguilar
GIDITIC, Universidad EAFIT
Universidad de Alcalá, Dept. Automática
Colombia, Venezuela, España
0000-0003-4194-6882

***Abstract*—Cattle breeding has been one of the most important industrial sectors in the world since it is related to food security and the survival of the human race. Management of the cattle fattening process is a fundamental procedure for cattle breeders because it allows them to make strategic decisions, such as timely treatment in case of any abnormality (e.g., weight gain in herds, in their paddocks). This article aims to present a management system for the cattle fattening process under a rotational grazing scheme, considering the health status of the animal and the pasture, which should diagnose weight loss or gain in bovines and recommend actions when is required. The diagnostic process is based on a fuzzy system that defines rules that characterize the diagnostic process to determine the current situation given an input. Furthermore, the fuzzy classifier optimizes its rules by means of genetic algorithms by modifying its membership functions, providing a more accurate system for diagnosis. On the other hand, the recommendation system is based on a classification model of pasture crops, in which the best pasture is recommended given the soil variables. We tested our proposal with experimental cases, with promising results. For the fuzzy classifier, the accuracy metrics are very good, with values of accuracy close to 100% and of *Area Under the Curve* close to 1. For the classification model were used several machine learning techniques, resulting in the best classifier the random forest technique, with an accuracy of 98.61%.**

***Index Terms*—Precision Livestock Farming, Rotational Grazing, Diagnostic System, Recommendation System**

I. Introduction

The livestock activity in the nations located in the tropics is becoming more and more demanding due to demographic growth, so this activity must constantly assume dynamics of transformation and adaptation. On the other hand, the intertropical climatic seasons are characterized by winter and summer, which condition the livestock activity in terms of animal feeding, so it is necessary to improve the feeding processes considering these aspects so that this activity generates greater volumes of profit. At the same time, the livestock industry presents a series of difficulties to detect failures in its production systems, which directly affects the ideal growth of livestock and the optimum production of paddocks. The problem lies in the inability to make decisions in real time, using data, of what may be happening in the cattle fattening process. Therefore, it is necessary to develop a management system for the cattle fattening process, with the objective of helping farmers make decisions to increase production efficiency.

Some works related to our proposal are the following. Palomino and Loza [1] designed a rotational grazing system for a high Andean dairy herd. The purpose of this project was to improve traditional grazing systems for dual-purpose high Andean dairy cattle. SAS Planet software was used for the design, and plans of the farm and areas for agriculture and housing purposes were plotted to determine the ideal number of paddocks and the optimal stocking rate.

Several authors [2], [3] outline the essential aspects for developing low-carbon livestock farming. They considered a set of effects related to climate change, and their impact on the soil-plant-animal relationship. They propose an intensive grazing system that guarantees low-emission livestock farming, which represents an option in the face of climate change. Garcia et al [4]. propose an approach to detect anomalies in the cattle fattening process. This approach uses the actual historical record of animal weight to identify whether animals have gained the appropriate weight over time. They compare several machine learning techniques (Decision Tree, Gradient Boosting, regression based on K-Nearest Neighbors and Random Forest) in the task of anomalous weight detection, using Mean Absolute Error as quality metrics.

On the other hand, regarding fuzzy classifier systems, Ramirez et al [5]. designed a fuzzy classifier system for the establishment of the functional states of a medical production system. For the design and tuning of the fuzzy classifier, the process data history was used to identify all functional states useful for process monitoring. The establishment of functional states from the fuzzy classification allows the programming of corrective actions in the process from the diagnosis performed. Finally, in [6] they applied fuzzy logic and genetic algorithms for the determination of treatments in malignant neoplastic diseases. The paper proposes an optimized fuzzy classifier using a hybrid genetic algorithm with a fuzzy clustering technique. They implement a prototype and evaluate it on synthetic treatment data against malignant neoplastic diseases. Finally, previous work has determined that machine learning is essential to provide self-management capability in a beef production farm [7].

This work aims to propose a management system for the cattle fattening process, which contains a system to diagnose the individual fattening progress, and the causes of cattle weight loss or gain, in rotational grazing, but in addition, it has a system to recommend the best pasture given the soil variables for each paddock of the farm. *The diagnostic system* is based on fuzzy theory, and uses rules to analyze the current cattle weight situation. In addition, the system optimizes the rules to fit the actual data of the cattle under supervision. On the other hand, *the recommendation system* is based on a classifier model built with machine learning techniques. Particularly, 3 algorithms were used: nearest neighbors, decision tree and random forest, resulting the last one with higher accuracy.

The organization of this work is as follows. Section 2 presents the theoretical framework used in this work. Section 3 shows the design of our cattle fattening process management system. Then, Section 4 describes the experiments conducted with the system to evaluate its quality. Next, Section 5 compares this work with previous work. Finally, Section 5 presents the conclusions.

II. Theoretical Framework

A. Fuzzy Classifier System

In general, fuzzy systems have proven to be very useful for representing the behavior or dynamics of systems by means of fuzzy rules. Traditionally, these systems are based on information provided by experts; however, in complex systems, the rules thus constructed did not allow an accurate simulation of the system [8]. The search for fuzzy systems that adapt to the dynamics of complex systems has led to the development of research on techniques for extracting fuzzy rules from input and output data [9]–[11]. Fuzzy classification algorithms represent one of the techniques for the development of adaptive fuzzy systems [12], [13].

A Fuzzy Classifier System is composed of two major components. A *rules subsystem* that allows classifying the input information. This subsystem defines the rules of type *If-Then* using fuzzy variables, and uses a fuzzy reasoner to infer a conclusion. The second component of a fuzzy classifier system is the *adaptive subsystem*. Traditionally, this subsystem is based on genetic algorithms (GA), which are inspired by biological evolution as a strategy for solving optimization problems [14]. In particular, the fuzzy classifier uses it to optimize the rules, and thus, its performance.

B. Classifier system

Our recommendation system is based on a classification model, which is previously trained with the information to be recommended, such that for a given input, it knows which recommendation to give.

The classification model is built using supervised machine learning techniques. Specifically, these models are built from a training process using a set of labeled data. Then, that trained model is used to predict/determine the class to which a new unknown input belongs. Thus, once trained, it is used to assign a label or class to a new data entry [15].

The three machine learning algorithms used in this work are as follows:

- Near neighbors: This algorithm is based on the idea that an entry is more similar to entries that are close to it in the feature space. [16].
- Decision Tree: This algorithm builds a tree that represents the decisions that must be made to arrive at a classification [17].
- Random Forest: This algorithm combines several decision trees to improve overall accuracy and reduce model variability [18].

C. Precision livestock farming

Precision livestock farming is one of the sectors that has incorporated the most innovations in recent years, thanks to the development of technologies aimed at improving the efficiency of livestock farms and the quality of animal products. These technologies make it possible to optimize resources, increase yields, control environmental impact and improve animal welfare. This requires monitoring animal health and production, measuring physiological and morphological biometric indicators, among other things [19].

In the case of diagnostic processes, the use of precision livestock farming allows the development of a friendly, non-intrusive monitoring technology, allowing interaction between experts and producers. In particular, diagnostic systems in precision livestock farming seek to reduce losses and improve productivity in the long term, through dynamic follow-up and monitoring of livestock. For example, a diagnostic system could evaluate quantitative information to characterize animals quickly and adequately, by studying individual and/or batches of animals. The system would interpret animal health and determine paddock performance to identify patterns in the production process, and from there, be able to diagnose possible causes of low or high fattening of cattle [4].

III. Design of the Cattle Fattening Process Management System

This section will discuss the design and operation of the fuzzy classifier-based diagnostic system and the classification model-based recommender system.

A. Diagnostic System Specification

The architecture of the diagnostic system is based on a fuzzy classifier, which can be seen in Figure 1. In this scheme, 2 blocks are observed, the first block, called *fuzzy rule design and optimization system*, and the second one is called the

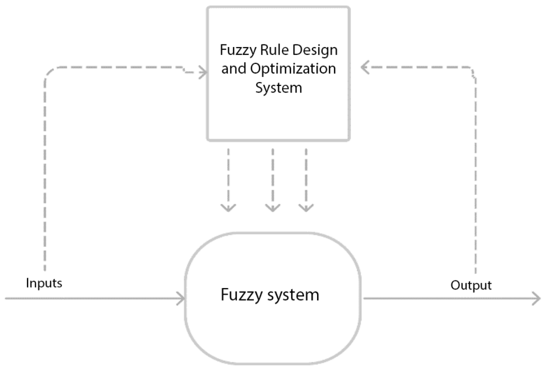


Fig. 1: Fuzzy classifier architecture.

fuzzy system. In Figures 2 and 3, respectively, both systems are detailed.

Figure 2 shows the fuzzy rule design system and optimization system. The rules are designed and created using a fuzzy clustering algorithm. The present work uses the *fuzzy-c-means* (FCM) algorithm, which allows fuzzy data clustering (each cluster determines a fuzzy set for the linguistic/fuzzy variable of interest) [1], [20]–[23]. On the other hand, a GA is used for optimization, whose task is to improve the definition of the fuzzy sets (it adapts the membership function of each fuzzy set).

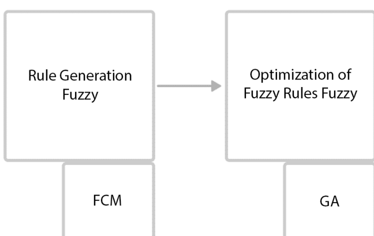


Fig. 2: Fuzzy Rule Design System and Optimization System

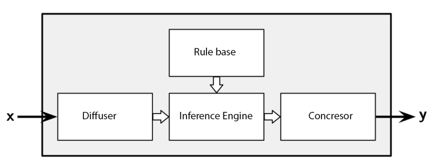


Fig. 3: Fuzzy system

Figure 3 shows the fuzzy system. Its components are the fuzzified input, the inference engine (fuzzy reasoner), the output generator (defuzzifier), and the knowledge base (the fuzzy rules), which is updated from time to time by the fuzzy rule optimization system shown in Figure 2.

B. Specification of the fuzzy system

1) Fuzzy system: The engine of this system is centered on the definition of linguistic variables that make up the diagnostic model. This model is constituted by a set of fuzzy rules that use linguistic variables. Thus, the fuzzy system represents the diagnostic model proposed in this work.

The first step in designing the fuzzy system consists of creating the labels (fuzzy sets) needed for each linguistic variable. The FCM algorithm allows defining the fuzzy sets from the membership matrix that it generates for each linguistic variable [1], [24]. For this purpose, FCM determines the fuzzy groups present in a linguistic variable (these will be the fuzzy sets), with the degree of membership of the values of the variable in each of them. The degrees of membership are in the range $[0, 100]$, and represent the partial membership in each fuzzy set (class) of each value of each linguistic variable. The linguistic variables, and their fuzzy sets, of our fuzzy diagnostic system, are described below.

The *linguistic variables* used are those linked to the fattening process. The input variables are the age, the paddock, and the differential of the final weight with the initial weight, and the output variable is the state of the animal.

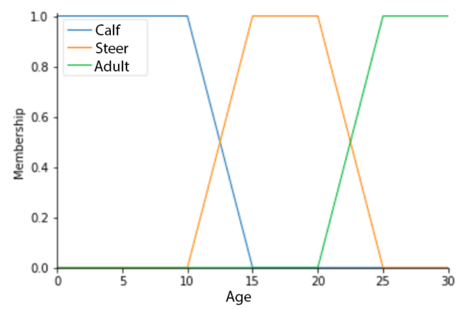


Fig. 4: Variable Age

- *Age:* its universe of discourse is $[0, 30]$, and represents months of life. Its fuzzy sets are $[calf, steer, adult]$. Figure 4 shows each fuzzy set.

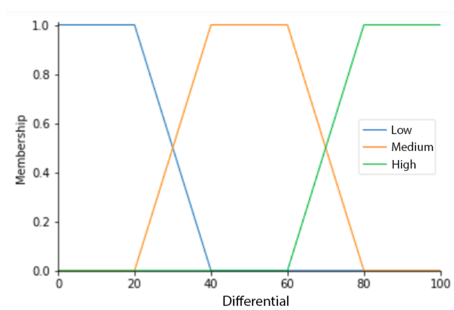


Fig. 5: Variable Diferencial

- *Differential:* its universe of discourse is $[0, 100]$, and represents the percentage of difference between the initial and current weight. Its fuzzy sets are $[low, medium, high]$. Figure 5 shows each fuzzy set.
- *Paddock*: its universe of discourse is $[0, 10]$, and represents the paddock yield score, characterized by forage quality and forage tolerance to the climatic situation. Its

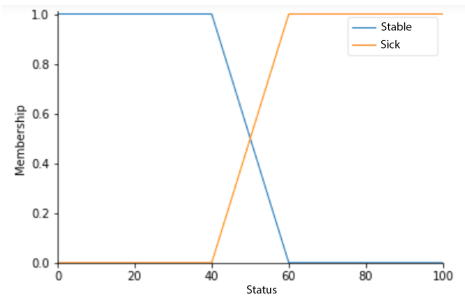


Fig. 6: Variable Paddock

fuzzy sets are $[bad, fair, good]$. Each fuzzy set is shown in Figure 6.

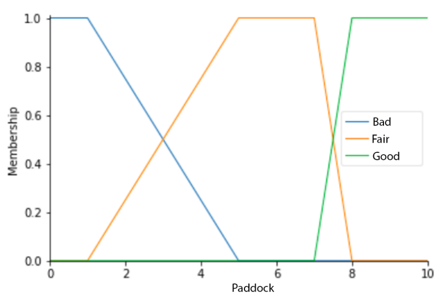


Fig. 7: Variable Animal Status

- *Animal Status*: its universe of discourse is $[0, 100]$, represented by a percentage to indicate how well the animal is doing. Its fuzzy sets are $[stable, sick]$. Figure 7 shows each fuzzy set.

2) Definition of diagnostic rules: The rules that model the system are of the form.

$IF < \ldots > THEN < \ldots >$, In the following tables, we define the set of rules.

There are two different types of fuzzy rules, depending on whether information on paddock performance is available. Thus, we have a group of rules with two antecedents (see Table I) or with three antecedents (see Table II). Particularly, in Table I the two antecedents are the variables differential and age; and in Table II the same, but now also the paddock variable (the values of it that directly influence animal fattening, such as fair/bad). An example of a rule for the latter case is:

If the age is calf and the differential is low, and the paddock is bad, then the animal's condition is sick.

These rules will be fitted/optimized to the input dataset using GAs. Specifically, the membership functions of the fuzzy sets will be those fitted to the input data.

TABLE I: Rule Base with 2 antecedents

		Diferencial		
		low	medium	high
Age	Calf	Sick	Sick/Stable	Stable
	Steer	Sick	Sick/Stable	Stable
	Adult	Sick/Stable	Sick/Stable	Stable

TABLE II: Rule Base with 3 antecedents

		Diferencial				
		low	medium	high		
Age	Calf	Sick	Sick/Stable	Stable	Fair/Bad	Paddock
	Steer	Sick	Sick/Stable	Stable		
	Adult	Sick/Stable	Sick/Stable	Stable		

C. Characteristics of the Genetic Algorithm

1) Specification and evolution of individuals: Structure of individuals: A factor of special interest in the design of the evolutionary process to optimize the fuzzy classifier is the representation scheme used to encode each of the possible solutions. For this research, the individual is defined as follows:

Chromosome = limits of the membership function of the fuzzy sets of each linguistic variable.

In this approach, each individual represents by itself a complete solution by encoding the fuzzy sets of input and output variables. Specifically, each gene represents a vertex of the membership function of the sets of each linguistic variable, assuming a trapezoidal shape for the definition of them.

The representation of each linguistic variable would be as follows (vertices of the membership functions of their fuzzy sets):

$$Age = [10, 15, 20, 25] = [e1, e2, e3, e4]$$
$$Differential = [20, 40, 60, 80] = [d1, d2, d3, d4]$$
$$Paddock = [1, 5, 7, 8] = [p1, p2, p3, p4]$$
$$Status = [40, 60] = [c1, c2]$$

The coding of the chromosome would be the concatenation of the description of the membership function of the fuzzy sets of each variable. On the other hand, since there are rules that do not contain the fuzzy variable paddock, there are 2 cases for chromosome coding, the first one does not contain the fuzzy sets of the fuzzy variable paddock, and the second case does. Thus, in the first case, considering the order of the genes, such that the antecedents (age and differential) are placed first and then the consequent (state), the individuals are as follows (see Figure 8):

In the second case of fuzzy rules, in which paddock performance is considered in the analysis, the antecedents would be age, differential and paddock, and the consequent state (see Figure 9).

e1	e2	e3	e4	d1	d2	d3	d4	c1	c2

Fig. 8: Chromosome for GA in case 1

e1	e2	e3	e4	d1	d2	d3	d4	p1	p2	p3	p4	c1	c2

Fig. 9: Chromosome for GA in case 2

2) Evolution of individuals: For the optimization of fuzzy rules with GA, an initial population of individuals must be randomly generated. The objective of the GA is to modify the fuzzy sets (their membership functions) in order to adapt the rules to the data. The evolutionary process is as follows.

- Individuals are selected for breeding using the following scheme: individuals are ordered from highest to lowest aptitude, and those with aptitude lower than 0.7 are discarded. The aptitude function is defined in the next section.
- Subsequently, genetic operators are used to generate new individuals. In particular, the crossover and mutation operators are used, which were implemented as follows:
  1) Crossing: two chromosomes and the cut point are chosen randomly, to perform the cross between the two.
  2) Mutation: a chromosome is chosen and one of its genes, chosen at random, is modified.
- In each iteration, a number of offspring equal to the initial population size is generated, and the worst individuals in the population are replaced by the best new ones, according to the aptitude function.

At the end of this, if the optimum is reached or the stop condition is met, the individual with the best value in the aptitude function is selected to update the vertices of the fuzzy sets of the fuzzy variables.

3) Aptitude Function: Aptitude function based on the calculation of the following measures was used:

Score: Measures the accuracy of an individual. In this case, as it is based on labeled data on diagnosis, the score establishes whether the consequent is true (the rule is triggered) when it should. It is calculated by the following equation:

$$Score = \frac{pv}{(pv + pf)} \tag{1}$$

Where pv is the number of true positives, and pf is the number of false positives. In particular, the score for each rule is calculated for the input dataset, and averaged.

Certainty or trigger degree of each rule: which is the activation level of a rule, calculated as the product of the membership functions of the antecedents. If the value is high, it means that the activation level of the rule is high.

Finally, the aptitude function of each individual is the average of the score of each rule multiplied by its degree of certainty. Thus, the evolutionary process, using this hybrid fitness function (score and degree of certainty), allows adjusting the fuzzy sets to fit the context data.

D. Specification of the Recommender System

In general, the recommendation system is based on a classifier model of livestock pasture crops. Particularly, the classification model is trained so that for a given input of soil variables it indicates the best pasture to consider. The following steps were considered for the construction of the model:

1) Dataset acquisition: Several datasets were used from the Kaggle Crop Dataset and the Pasto certo page www.pastocerto.com/ on pasture information. However, since information about soil physicochemical characteristics in relation to grass is very scarce, it was decided to take data from other crops similar to the different types of pastures, with the objective of having a robust dataset to train and evaluate the model. In addition to the pasture dataset, cattle were also required. Thus, the data sets used were:

- Soil nutrient content and environmental data set: This data set contains information on the main soil physicochemical variables (soil nutrients) and some environmental characteristics (maximum and minimum rainfall, maximum and minimum temperature) for 22 types of crops.
- Cattle weight gain dataset: This dataset contains information on cattle weight gain (kg) for 18 pasture types, in dry and rainy seasons.

The specific variables used from the data sets were as follows:

- Temperature: Monthly temperature (in °C).
- Moisture: relative amount of water contained in a saturated soil after 48 hours of drainage.
- Weight gain—water: gain in grams of cattle weight per day in the rainy season.
- Weight gain—dry: gain in grams of cattle weight per day in the dry season.
- Potassium: helps the plant to make more efficient use of water, it is absorbed in ionic form (K+).
- Nitrogen: It is found in two different forms: organic and chemical.
- pH: A measure of the acidity (low pH = acidic) or alkalinity (high pH = basic or alkaline) of the medium.
- Rainfall: potential capacity of rainfall to erode soils.
- Phosphorus: An essential nutrient for crops, and its lack can significantly limit crop growth and yield.

2) Data preprocessing: Since several datasets are taken for the construction of the model, the data were unified in one dataset. Thus the new dataset contains information on the physicochemical characteristics of the soil in various crops, with the weight gains of cattle in pastures. To build the dataset, modifications were made to the soil nutrient dataset, changing the crop names to pasture types in order to combine them with the cattle weight gain dataset.

On the other hand, data preprocessing also involved substituting null values so as not to affect the overall prediction, as well as treating missing values or outliers to eliminate them.

3) Construction of the Classification Model: In this phase, supervised learning techniques were used to build the classification model. Three techniques were used to build the classifier model: K-nearest neighbors, Decision Tree and Random Forest.

The techniques that were used are because they are widely used and known in machine learning to solve classification problems. Each of them has different strengths and weaknesses, so it is common to evaluate them beforehand to determine which one is the best fit for a specific classification task.

The results were analyzed and compared in relation to cross-validated pasture classification accuracy, as shown in Table III. Taking as the evaluation metric the accuracy of the models, it is obtained that random forest is the best model, with a substantially high accuracy, generating an efficient classification of the categories created, with an average of 98.61% accuracy for the classes generated. The other techniques are relatively close, but the recommendation system will be defined using the classification model built with the random forest.

TABLE III: Comparison between models

Algorithms	**Accuracy**
Nearby neighbors	98.05%
Decision tree	96.66%
Random forest	98.61%

IV. Experimentation

In this section, the behavior of our system is analyzed. For this purpose, an experimental protocol is developed describing the context where the management system is tested, and its performance is evaluated by means of quality metrics.

A. Experimental Context

Simulator For the experiments, a livestock simulator is used (see https://github.com/devraxielh/Simulador_Ganadero), which emulates all the climatic and soil conditions of a farm, as well as the fattening behavior of the cattle. In particular, the simulator generates the values of the variables that describe the land to which rotational grazing is applied, and of the cattle population being fattened. In this way, the simulator randomly generates the data that feed the fuzzy classifier system. The data of interest provided by the simulator are the following: number of animals, age of each animal, the initial weight of each animal, the evolution of the weight of each animal, number of paddocks, forage of each paddock, the climatic season of each paddock, and forage tolerance to hot climates.

1) Quality Metrics: The metrics for assessing the quality of the diagnostic and recommendation systems aim to estimate the accuracy of the models on test datasets (different from those used in the adaptive process of the rules, or training for the classification model). These are:

- Accuracy (score): It is the percentage of correct hits given by the adaptive system.
- Certainty degree: This is the value obtained by multiplying the membership functions of the antecedents to determine the degree of triggering of each rule and to deduce which rule is triggered more than another.
- ROC (Receiver Operating Characteristic) curve: This is a graph that shows the sensitivity and specificity of a model:
  - Sensitivity: The probability that the model predicts a positive outcome for an observation, when in fact the outcome is positive.
  - Specificity: The probability that the model predicts a negative outcome for an observation, when in fact the outcome is negative.
- Area Under the Curve (AUC): It is a numerical value given by the area of the curve generated by the ROC plot. The larger the area covered, the better the machine learning models will be at distinguishing the given classes. The ideal value for AUC is 1.

B. Diagnostic System Test Scenarios

This section presents the case studies that will be analyzed by the diagnostic system. Each case study represents situations where climatic situations that may or may not affect the paddock are prevalent, and animal health situations are also considered. The following is a description of the case studies considered.

1) Case study 1:

a) Objective: The first case of study is the optimal case, where all animals are stable, and the paddock is suitable for use. The rules that will be activated are those of the group of 2 antecedents, since the case is ideal, then the performance of the paddock will be good and will not negatively affect the animal fattening.

b) System input: Table IV presents a partial view of the historical cattle fattening data (the first 4 columns). It shows the input variables (age, initial weight and final weight), but also has an identifier for each animal. On the other hand, the initial weight indicates the weight at which they enter the paddock, and the final weight represents the weight at which they leave the paddock. Considering that daily animal gain varies on average from 300 to 500 grams per day, it can be said that fattening is normal if after 30 days (time spent in each paddock) they gain approximately 8 to 15 kilos. This determines that the weight differential is within the range, and therefore it is an ideal case.

Table IV shows in the column degree of membership of each data to the stable label group; hence, the prediction column shows the binary label 1 (as stable). These two columns represent the current prediction made by the classification system for the input data set. Table V shows the paddock input data. It shows the occupied paddock, the name of the forage used in the paddock and the weather situation. Table V shows the forage column, which represents the score the paddock has on a range of [0,10], with 10 being the highest score. This data is obtained depending on the tolerance of the forage to the climatic situation in which it is found. The status column represents the performance of the paddock according to the fuzzy interpretation to categorize paddock performance.

TABLE IV: Representation of livestock data and classification of animal status in experiment 1

Age (months)	Weight initial(kg)	Weight end(kg)	ID animal	Prediction	Grades membership
19	373.0205	384.4683	V1	1	0.9171
29	389.0205	400.4683	V2	1	0.9171
23	368.0205	379.4683	V3	1	0.9171
18	402.0205	413.4683	V4	1	0.9171
22	385.0205	396.4683	V5	1	0.9171
19	404.0205	415.4683	V6	1	0.9171
22	369.0205	380.4683	V7	1	0.9171
22	407.0205	418.4683	V8	1	0.9171
19	400.0205	411.4683	V9	1	0.9171
26	418.0203	427.1916	V10	1	0.9171

TABLE V: Paddock performance in experiment 1

Name	Climate	Pasture	Forage	Status
Humidicola comum Tuly o Quicuio da Amazönia	Rain	6	7	Good

c) Analysis of results: As could be seen, with the existing fuzzy rules, the fuzzy classifier system infers that all animals are stable for those input data. Once several iterations of the fuzzy system have been performed, the adaptive system of the rules is invoked (see the following figures showing the adaptation of the rules to the data).

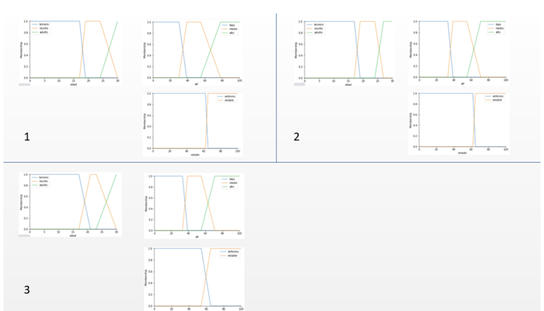


Fig. 10: Adaptation of membership functions

It can be seen in Figure 10 how the fuzzy sets of each fuzzy variable are fitted using the dataset in table IV. Each represents the best individual for the first 3 iterations of the GA. For example, we see in Figure 11.1 the fuzzy sets of the antecedent and consequent of the best individual for the first iteration. The fuzzy variable of the paddock is not activated, since the performance of the paddock is good. The fuzzy rules that are activated by these settings, using the dataset from table **??**, are shown in Figure 11.1. Figure 11. 2 shows the best individual of the next iteration, and its rule base, in Figure 11.2. Finally, Figure 11.3 shows the best individual when the GA is stopped, either because the stopping condition is satisfied or because it found the optimum. The final rule base shown in Figure 11.3 are the necessary system rules for the dataset in Table IV. The individual in Figure 11.3 represents the potential solution, and by comparing the final activated rules (Figure 11.3) with the dataset shown in the IV box labeling all stable animals, we see that consistently the activated rules have in the consequent only the label stable.

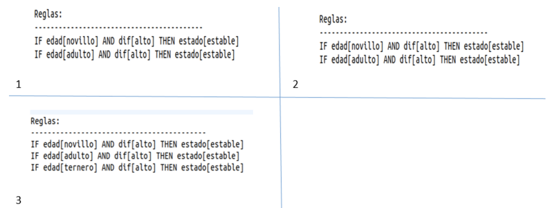


Fig. 11: Evolution of the activated rules in case 1

TABLE VI: Case 1 metrics.

Metrics	Values
Accuracy	100%
Certainty of rules	R1:0.85, R2: 0.88,R3: 0.89
AUC	1

Finally, with the final fuzzy rules (see Figure 11) we proceed to diagnose using the test dataset. With the result of the inference process for each individual, we proceed to calculate the averages of the system quality metrics (see Table VI and Figure 12).

The ROC curve determines the performance of the model, and allows us to calculate the AUC, which would be the area under the curve. We note in table VI that AUC is 1 and the precision is 100%. Recall that the certainty of each rule (Ri) is determined by its degree of triggering, which varies between 0 and 1. We see that all 3 rules have a high level of triggering (greater than or equal to 0.85). The above results tell us that the model is robust and performs a perfect classification.

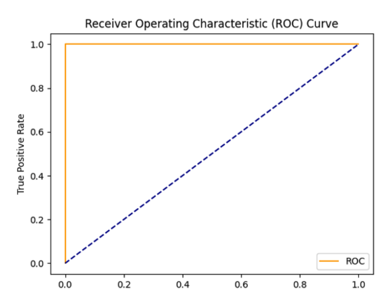


Fig. 12: ROC curve case 1

Case study 2

d) Objective: The second case study presents a variation, where at least one animal is labeled sick, and paddock performance is good, despite the weather season. The rules that will be activated are those of 2 antecedents, since the paddock performance will be good and will not negatively affect animal fattening.

e) System input: The input data are as follows:

TABLE VII: Representation of livestock data and classification of animal status in experiment 2

Age (months)	Age initial(kg)	Age end(kg)	ID animal	Prediction	Grade membership
19	332.0	334.5431	V1	0	0.2543
29	348.0	356.5431	V2	1	0.8543
23	327.0	335.5431	V3	1	0.8543
18	361.0	369.5431	V4	1	0.8543
22	344.0	352.5431	V5	1	0.8543
19	363.0	371.5431	V6	1	0.8543
22	328.0	336.5431	V7	1	0.8543
22	366.0	374.5431	V8	1	0.8543
19	359.0	367.5431	V9	1	0.8543
26	380.0	388.1348	V10	1	0.8134

Table VIII contains the data used for paddock entry. It shows the occupied paddock, the name of the forage used in the paddock, and the weather condition.

TABLE VIII: Paddock performance in experiment 2.

Name	Climate	Pasture	Forage	Status
Humidicola comum Tuly o Quicuio da Amazönia	Seco	1	7	Bueno

The same table VIII shows the forage column, which represents paddock quality. The status column shows the paddock yield.

f) Analysis of results: With the existing fuzzy rules, we saw that the fuzzy classifier system infers that only one animal is sick for that given input (see table VII). Once several iterations of the fuzzy system have been performed, the adaptive system of rules is invoked.

The evolution of the best individual through the generations of the GA is observed in Figure 14. The rules activated using the best individual of the 1st generation (see Figure 14.1) and the training dataset, are shown in Figure 13.1, but these rules do not satisfy the system, since it only has rules with stable consequent. In the 2nd iteration, in Figure 14.2, its activated rules with the training dataset can be seen in Figure 13.2. Already in that iteration, an improvement in the rules is noticed because the sick state appears in the consequent.

Finally, in the 4th iteration, the stop condition is reached, and its best individual represents the final solution (see Figure 14.4). In this case, the final rules (see Figure 13.4) are the ones needed for the diagnostic system for the dataset shown in table VII.

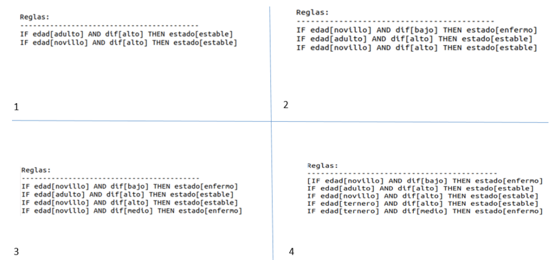


Fig. 13: Evolution of the activated rules in experiment 2.

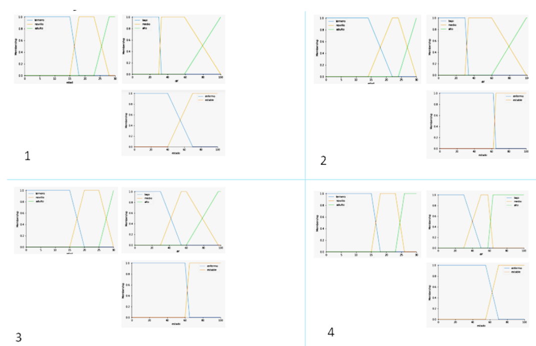


Fig. 14: Adaptation of membership functions.

Finally, with the final fuzzy rules (see Figure 13.4), the diagnostic system is tested using the test data. With the result of the inference process, the quality metrics of the system are calculated (see table IX and Figure 15).

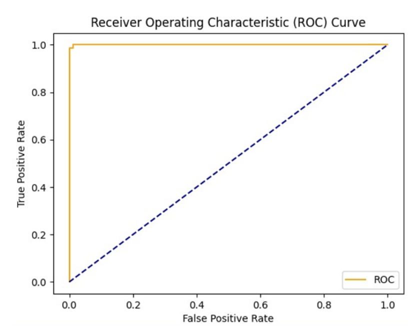


Fig. 15: ROC curve of case 2.

TABLE IX: Case 2 metrics.

Metrics	**Values**
Accuracy	97%
Rule certainty	R1:0.85, R2:0.88,R3:0.90,R4:0.85,R5:0.85
AUC	1

We note in table IX that AUC is 0.98, and the precision 97%. On the other hand, the certainty of each final fuzzy rule is equal or greater than 0.85. All of the above tells us that the model is robust, performs a good diagnosis, and its rules are adequate for the input data.

2) Case study 3:

a) Objective: The third case study presents a variation, which consists in analyzing the animal fattening, taking into account the paddock performance when being regular. The rules that will be activated are those of 3 antecedents, in this case, the paddock variable is activated because it has a negative effect on animal fattening.

b) Input data is as follows:

TABLE X: Representation of livestock data and classification of Animal status in experiment 3.

Age (months)	Weight initial(kg)	Weight end(kg)	ID animal	Prediction	Grade membership
21	434.8377	438.0388	V1	0	0.5201
29	450.8377	456.0388	V2	1	0.5201
23	429.8377	435.0388	V3	1	0.5201
22	463.8377	469.0388	V4	1	0.5201
22	446.8377	452.0388	V5	1	0.5201
22	465.8377	471.0388	V6	1	0.5201
22	430.8377	436.0388	V7	1	0.5201
27	443.8223	448.0238	V11	0	0.4201
23	451.8223	456.0238	V12	0	0.4201
27	448.8223	453.0238	V13	0	0.4201

Table X presents a partial view of the historical cattle fattening data. Based on what was said for the previous scenarios, we note that there are several animals that are outside the ideal range (see first 4 columns). We look for the fuzzy classifier system to identify them, now also, using the paddock variable.

Using our classifier system, we see in table X, in the column degree of membership, to the stable label in low in some cases (it determines that value). This means that they belong very little to the stable label (that is why we see in the prediction column their label with the value 0).

TABLE XI: Experiment 3 paddock data .

Name	Climate	Paddock
BRS Zuri	Seca	1

Table XI contains the data used for paddock entry. It shows the occupied paddock, the name of the forage used in the paddock, and the weather condition.

Figure 16 shows the 4 fuzzy variables (now with the paddock), and how their fuzzy sets are being adjusted. Figure 16.1 shows the best individual of the first iteration and the rules activated with the training dataset in Figure 17.2, and so on through the iterations. Figure 16.3 shows the final solution of the GA when it converges (individual with the highest value in

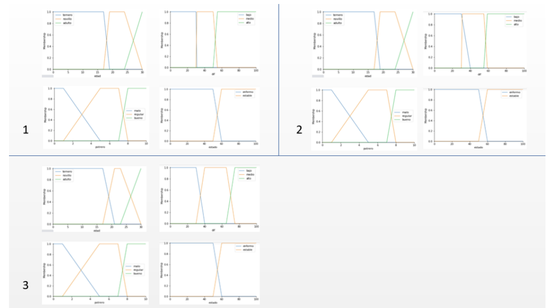


Fig. 16: Adaptation of membership functions.

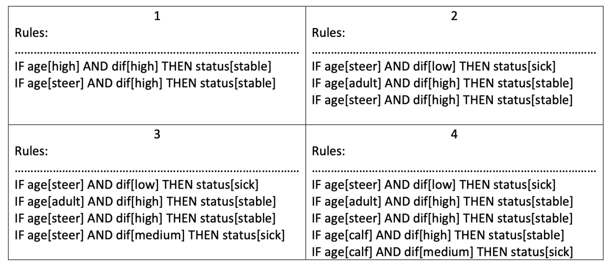


Fig. 17: Evolution of the rules that are activated in case 3.

the fitness function), and Figure 17.3 shows the rules activated with the training dataset. This will be the final fuzzy rule base. There are a larger number of rules, as more rules are needed to control the system.

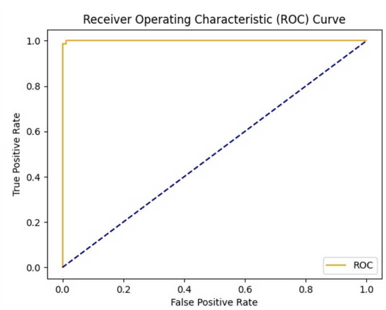


Fig. 18: ROC curve in case 3.

TABLE XII: Metrics of case 3.

Metrics	**Values**
Accuracy	95%
Certainty of rules	R1:0.88, R2:0.75, R3:0.89, R4:0.85, R5:0.85, R6:0.75
AUC	1

Finally, with the final fuzzy rules, the system is tested using the test data, and the quality metrics are calculated (see Table XII and Figure 18).

C. Recommendation System Test Scenario

This system is activated once situations such as case 3 are detected, where the weight loss is caused by the grass in the paddock. Case 3 indicates that it is important to take action, as the grass in the paddock is not performing. This situation has a direct impact on cattle fattening and, ultimately, on the profitability and success of the cattle business. A pasture that is not performing adequately will not provide the right amount of feed for the cattle, resulting in reduced beef production.

In particular, the recommendation system is able to recommend, with the environmental parameters and soil characteristics of the pasture, the best type of grass to use. Specifically, in case 3, paddock 1 is planted with BRS Zuri grass and the climate is dry. Immediately, a soil survey is done to establish all the input variables to the classifier model (see Table XIII).

TABLE XIII: Input to the classifier model.

Variable	Value
Nitrogen	95
Phosphorus	42
Potassium	43
Temperature	20.87974371
Humidity	82.00274423
pH	6.502985292000001
Rainfall	202.9355362
Water weight gain	600 a 640 gr/day
Dry weight gain	420 gr/day

These data are fed to the classifier model trained with random forests, and the recommendation given by our system is that the best pasture for paddock 1 should be BRS Paiaguás (see Table XIV). Thus, the recommendation system indicates the best type of grass to use in that paddock to improve animal fattening.

TABLE XIV: Recommendation for paddock 1 in case 3.

Name	Climate	Paddock
BRS Paiaguás	Seca	1

V. Comparison with other studies

To compare this work with other similar works, we proceeded to define four criteria, which are:

- $Cri1$: The work uses non-intrusive schemes for the supervision of the animal fattening process.
- $Cri2$: The work uses intelligent techniques.
- $Cri3$: The work uses machine learning to improve the monitoring process.
- $Cri4$: The work jointly analyzes grazing and animal welfare.

These criteria are relevant in precision livestock farming because they fulfill 2 important aspects, which are the use of automatic technologies of industry 4.0, and strive for animal welfare to improve livestock production. Comparison with previous work is shown below in Table **??**.

TABLE XV: Comparison with other works.

	Cri1	Cri2	Cri3	Cri4
[2]	✓	✗	✓	✗
[3]	✓	✗	✓	✗
[8]	✗	✓	✓	✗
[5]	✗	✓	✓	✗
This work	✓	✓	✓	✓

The first criterion is met by [2] and [3], since both use computerized methods and models for the supervision process, specifically for diagnosis. The second criterion is met by [8] and [5], since they use fuzzy logic in their architecture, in one case to find malignant diseases, and in the other for the generation of a medicinal production system. The third criterion is met by all of them, since they use machine learning for different things, for example, to process acoustic information from animal chewing [2], study the environment of livestock [3], detect diseases [8], or establish the best medicinal production system [5]. The fourth characteristic is only fulfilled by our work since it is aimed at precision livestock farming in a rotational grazing environment that seeks animal welfare.

This work meets all the criteria that point to a management model of an animal fattening process, it even goes much further because, in addition to diagnosing, it recommends concrete actions. Thus, it ensures animal welfare through an adaptive fuzzy diagnostic system based on evolutionary learning (using GAs), and a recommendation system based on a classifier model to estimate the best pasture for a given paddock, all essential in precision livestock farming to ensure animal health, and make the best use of livestock in rotational grazing.

VI. Conclusions

This research presented the development of a system that manages the animal fattening process in rotational paddocks. The system is composed of a fuzzy diagnostic system designed by integrating FCM, a fuzzy reasoner and GAs, in order to obtain a diagnostic model adaptive to the environment data. In addition, the management system is composed of a recommendation system based on a classifier model designed to estimate the best grass given the soil characteristics.

The proposed diagnostic system allows adapting a set of rules to diagnose the fattening process of a batch of cattle using a fuzzy inference process. This fuzzy system is characterized by its flexibility and its tolerance to inaccuracy [25], since it can perform approximate reasoning using information from the environment (the input data to the fuzzy system). On the other hand, the recommendation system uses a grass classifier model that learns from the environmental and physical-chemical variables of the soil to determine what kind of grass to use in each case.

The effectiveness of the proposed method has been demonstrated by several experimental cases, with results that are promising. The accuracy metrics have a high value, indicating that the diagnostic system manages to train quite well using the process data, with a low error rate in terms of false positives. On the other hand, certainty metrics are above 0.75, indicating that the activation levels of the updated rules are high, as they are adapted to the data (they define the usefulness of those rules). Finally, AUC is close to 1, which says that the classifier has a very low margin of error, and its level of error is almost zero. All these values confirm a high credibility in the diagnosis of the system. Likewise, the classification model has very high accuracy values (98.61%) that speak of its quality to determine the best grass for a given environmental context.

This proposal is an effective alternative to be applied in precision livestock farming, specifically in rotational grazing to diagnose cattle diseases, either by disease or by paddock performance. However, among the limitations of this work are that the system is only applicable to rotational grazing, the cattle population must be from the tropics, and only 2 climatic seasons (summer and winter) are evaluated. Thus, this system can only be used in contexts with the following characteristics: on farms that use rotational grazing, regardless of the size of the cattle herds, and that the climatic seasons are summer and winter, understanding that in the tropics, winter is a rainy season and summer is a dry season.

One of the future works is to test this management system as an autonomous cycle of data analysis tasks [26] to automate the monitoring of the animal fattening process in the framework of precision livestock farming. Another future work is to extend the process of rule adaptation (which is currently only based on fitting to trapezoidal type membership functions). For example, allowing the adaptation of fuzzy sets with other membership functions (e.g., Gaussian), or even the possibility of making changes in the variables used in the rule antecedent. Along the same lines, other possible extensions are to use more context variables (e.g., explicitly the weather), to allow the number of fuzzy sets in the fuzzy variables to vary (e.g., more or fewer states to characterize the fattening process), among other improvements. Other future work on the pasture classifier model is to use real data on the soil physico-chemical variables associated with the pasture crop for each paddock.

Acknowledgment

All authors would like to thank the Vice Rector's Office for Discovery and Creation, Universidad EAFIT, for their support in this research. Likewise, the authors would like to thank the HAMADI 4.0 project, code 22-STIC-06, of the STIC-AmSud regional program, for supporting this work.

References

[1] F. E. Palominos, H. M. Díaz, S. K. Palominos, and L. R. Cañete, "Relación entre los procedimientos de selección a la educación superior y el desempeño académico de los estudiantes con base en una clasificación mediante conjuntos difusos," *Formación universitaria*, vol. 11, no. 1, pp. 45–52, 2018.

[2] J. O. Chelotti, *Sistema de adquisición y análisis de información acústica para ganadería de precisión*. PhD thesis, 2018.

[3] Y. Qiao, M. Truman, and S. Sukkarieh, "Cattle segmentation and contour extraction based on mask r-cnn for precision livestock farming," *Computers and Electronics in Agriculture*, vol. 165, p. 104958, 2019.

[4] R. García, J. Aguilar, M. Toro, and M. Jiménez, "Weight-identification model of cattle using machine-learning techniques for anomaly detection," in *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 01–07, 2021.

[5] J. A. Ramírez Urrego, H. O. Sarmiento Maldonado, and J. M. López Lezama, "Diseño de un clasificador difuso para el establecimiento de los estados funcionales de un sistema de producción de aire medicinal," 2017.

[6] M. León, B. L. de Lameda, C. Lameda, J. G. Chacon, M. S. Martínez, J. Rojas, J. Contreras-Velásquez, M. Graterol-Rivas, S. Wilches Durán, M. Aguirre, *et al.*, "Aplicación de lógica difusa y algoritmos genéticos para clasificación de tratamientos contra enfermedades neoplásicas malignas," *Archivos Venezolanos de Farmacología y Terapéutica*, vol. 35, no. 2, pp. 36–41, 2016.

[7] R. García, J. Aguilar, M. Toro, N. Pérez, A. Pinto, and P. Rodríguez, "Autonomic computing in a beef-production process for precision livestock farming," *Journal of Industrial Information Integration*, vol. 31, p. 100425, 2023.

[8] G. Gabriel Elías Chanchí, M. Luz Marina Sierra, and M. Wilmar Yesid Campo, "Application of fuzzy logic in the implementation of evaluation rubrics in the university context [aplicación de la lógica difusa en la implementación de rúbricas de evaluación en el contexto universitario]," *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, vol. 2021, no. E42, pp. 164–177, 2021.

[9] M. Cerrada, J. Aguilar, E. Colina, and A. Titli, "Dynamical membership functions: an approach for adaptive fuzzy modelling," *Fuzzy Sets and Systems*, vol. 152, no. 3, pp. 513–533, 2005.

[10] J. Aguilar, "A fuzzy cognitive map based on the random neural model," in *Engineering of Intelligent Systems* (L. Monostori, J. Váncza, and M. Ali, eds.), (Berlin, Heidelberg), pp. 333–338, Springer Berlin Heidelberg, 2001.

[11] H. Sánchez, J. Aguilar, O. Terán, and J. Gutiérrez de Mesa, "Modeling the process of shaping the public opinion through multilevel fuzzy cognitive maps," *Applied Soft Computing*, vol. 85, p. 105756, 2019.

[12] A. C. Martínez and A. D. A. Armenteros, "La lógica difusa en la modelización del riesgo operacional. una solución desde la inteligencia artificial en la banca cubana fuzzy logic to model operational risk: A solution based on artificial intelligence in the cuban banking," *COFIN-HABANA*, no. 2, pp. 122–135, 2018.

[13] M. Cerrada and J. Aguilar, "A fuzzy classifier system for fault tolerance," *Revista Técnica de la Facultad de Ingeniería. Universidad del Zulia*, vol. 23, 1 1.

[14] C. A. Riveros Varela, M. A. Melgarejo Rey, A. Riveros Varela, and L. D. Alvarado Nieto, "Sistema difuso evolutivo para la predicción del nivel de contaminación del aire por material particulado: Caso puente aranda (bogotá),"

[15] P. L. Lanzi, "Learning classifier systems: then and now," *Evolutionary Intelligence*, vol. 1, pp. 63–82, 2008.

[16] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "Knn model-based approach in classification," in *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings*, pp. 986–996, Springer, 2003.

[17] S. B. Kotsiantis, "Decision trees: a recent overview," *Artificial Intelligence Review*, vol. 39, pp. 261–283, 2013.

[18] A. Cutler, D. R. Cutler, and J. R. Stevens, "Random forests," *Ensemble machine learning: Methods and applications*, pp. 157–175, 2012.

[19] R. García, J. Aguilar, M. Toro, A. Pinto, and P. Rodríguez, "A systematic literature review on the use of machine learning in precision livestock farming," *Computers and Electronics in Agriculture*, vol. 179, p. 105826, 2020.

[20] P. Melin, F. Olivas, O. Castillo, F. Valdez, J. Soria, and M. Valdez, "Optimal design of fuzzy classification systems using pso with dynamic parameter adaptation through fuzzy logic," *Expert Systems with Applications*, vol. 40, no. 8, pp. 3196–3206, 2013.

[21] I. S. Masich, M. A. Kulachenko, P. S. Stanimirović, A. M. Popov, E. M. Tovbis, A. A. Stupina, and L. A. Kazakovtsev, "Formation of fuzzy patterns in logical analysis of data using a multi-criteria genetic algorithm," *Symmetry*, vol. 14, no. 3, p. 600, 2022.

[22] A. K. Dubey, U. Gupta, and S. Jain, "Comparative study of k-means and fuzzy c-means algorithms on the breast cancer data," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 8, no. 1, pp. 18–29, 2018.

[23] J. Cózar, J. A. Gámez, *et al.*, "Learning compact zero-order tsk fuzzy rule-based systems for high-dimensional problems using an apriori+ local search approach," *Information Sciences*, vol. 433, pp. 1–16, 2018.

[24] C. Ramos-Palencia, D. Mújica-Vargas, and M. Mejía-Lavalle, "Análisis comparativo de la calidad de agrupamiento del algoritmo intuitionistic fuzzy c-means.," *Res. Comput. Sci.*, vol. 149, no. 8, pp. 365–378, 2020.

[25] J. Cabrera, "Classical logic and diffuse logic: facets that characterize them," *Estoa*, vol. 2, pp. 91–101, 05 2013.

[26] J. Aguilar, O. Buendia, A. Pinto, and J. Gutiérrez, "Social learning analytics for determining learning styles in a smart classroom," *Interactive Learning Environments*, vol. 30, no. 2, pp. 245–261, 2022.

Appendix H

Using meta-learning in a cattle weight identification system for anomaly detection

A meta-learning approach in a cattle weight identification system for anomaly detection

Rodrigo García[1,2,3], Jose Aguilar[1,4,5]

[1]GIDITIC, Universidad EAFIT, Medellín, Colombia.

[2]Faculty of Engineering, Universidad de Córdoba, Montería, Colombia.

[3]Faculty of Engineering, Universidad del Sinú, Montería, Colombia.

[4]CEMISID, Universidad de Los Andes, Merida, Venezuela.

[5]IMDEA Networks Institute, Leganés, Madrid, Spain.

rjgarciah@eafit.edu.co, aguilar@ula.ve

Abstract

Weighing management in cattle farming is important for farmers, as it allows them to accurately monitor the growth and development of their animals. It is also a valuable tool that allows farmers to maximize the production and welfare of their animals. However, it is difficult for the farmer to detect if the herd of animals being weighed is gaining the ideal weight for a given breed and age. In addition, normally, when a new breed of cattle is introduced to a farm, there is very little data. This article proposes a meta-learning framework (MTL) for identification models used in the fattening process of animals to detect anomalies in cattle weight. The proposed MTL framework has a knowledge base of Meta-Models on Identification models based on machine learning techniques, which is used to select the identification model to use when a new breed of cattle arrives on the farm. This knowledge base is updated, either because a previous identification model has been successfully adapted to the new breed, or a new identification model has had to be generated, allowing the framework to continuously improve its performance over time. Particularly, this article presents in detail the process of adaptation of the previous identification models to new breeds carried out by our MTL framework. Besides, to test our approach, a case study is presented, using records of animals raised and fattened at the "El Rosario" farm, located in the municipality of Monteria (Córdoba-Colombia). The results are very encouraging in terms of the ability of our framework to adapt the identification models to different possible scenarios in the process of detecting anomalous weights. In general, the identification models generated with our proposal had an R^2 of 90.8%, which suggests that the models can explain the variability observed in the data.

keywords: Meta-Learning, Identification Models, Artificial Intelligence, Precision Livestock Farming, Rotational Grazing, Beef Production, Anomaly detection

1 Introduction

A bovine's weight can be an indicator of its health and well-being. Cattle that are abnormally underweight may have health problems, such as disease or nutritional deficiencies [1].

On the other hand, cattle that are abnormally overweight may have health problems related to obesity, which can cause problems in the joints or the cardiovascular system [2]. Thus, detecting abnormalities in cattle weight is a major issue in the cattle industry [3]. Besides, good cattle weight management can improve efficiency in Beef Production (BP), which can translate into increased profitability for cattle producers.

Precision livestock farming (PLF) offers farmers a real-time monitoring and management system. PLF can provide a real-time warning when something goes wrong so that the farmer can take immediate action to solve the problem [4]. For this, it requires systems that allow it to carry out the identification process of abnormal situations. Particularly, identification techniques propose approximate models of a real system, based on linguistic or mathematical expressions or an algorithm [5]. System identification has had an important development [6], but many problems remain. One of these problems is the definition of models for control and yield adjustment in real-time [6].

On the other hand, Machine learning (ML) is a very useful tool for cattle farming [7]. It can be used to improve herd selection, and herd rotation management, among other things. Also, ML models can provide information about the health and performance of cattle, as well as the quality and quantity of food and water that must be supplied. ML can too help identify diseases early, allowing for faster treatment and better monitoring of a BP [8]. All these possible applications help livestock producers to make better decisions in the management of their BP [9]. Finally, it is important to emphasize that within the identification techniques there are those based on ML, which have produced very interesting results in different contexts [3].

Although ML can be a valuable tool in cattle ranching, it also presents some challenges [7]. One of the main challenges is the availability of high-quality data, as ML requires a large amount of data to operate efficiently. In this regard, MTL, also known as "learning about learning", is an ML paradigm that is used to improve the ability of an ML-based model to adapt to the context (new datasets, etc.) [10]. For example, MTL allows an ML model to "learn how to learn" from new datasets, which may even be very small [11]. In general, MTL allows an Automated-Machine-Learning (AutoML) process for the automatic selection, composition, and parameterization of ML models, to achieve optimal performance on a given task [12–14].

Specifically, in the context of cattle weight anomaly detection, MTL can be an important tool for the definition of detecting anomalies models with a limited amount of data. In this paper, we present a novel MTL framework, with the aim of automating the selection and/or parameterization of ML-based models of cattle weight identification. Our approach proposed in our work is a novel approach to meta-learning in the context of livestock weight anomaly detection. Our proposal is innovative in that it is based on the construction and continuous adaptation of metamodels based on the results and prediction quality of individual models (in this case, detection of anomalies). These metamodels are not predefined but are generated in real-time based on the ML models to be adapted and their quality metrics. Thus, our metamodels are built according to the context in where will be used the MTL framework. In particular, the proposed solution seeks to improve anomaly detection when little data exists for a specific cattle breed. The main contributions of this work are:

- The proposition of the first MTL architecture for livestock weight anomaly detection in the livestock industry, which can be successfully adapted to different scenarios in livestock production.

- The definition of an MTL framework that uses a meta-model knowledge base of identification models based on machine learning techniques, which allows the selection of the appropriate identification model for each breed of cattle on the farm.

- The characterization of an MTL approach for precision livestock farming, which can be adapted to different contexts of production for anomaly detection, among other applications.

The remainder of this paper is structured as follows: Section 2 introduces the related work to this work. Section 3 describes our MTL framework and Section 4 its instantiation in a case study. Furthermore, Section 5 shows an analysis and discussion of the results, and Section 6, a comparison with other works. Finally, Section 7 presents the conclusions and future works.

2 Related Work

Regarding ML for anomaly detection, considerable progress has been made in the use of tools for routine monitoring and collection of animal information in BP [15]. However, although there are many applications of ML in livestock, the use of meta-learning in this industry is new, and there are no works about the utilization of MTL specifically to improve detection of cattle weight anomalies during rotational grazing. For this reason, in this section, we present some works on anomaly detection using MTL in other contexts, and on anomaly detection in PLF.

2.1 Anomaly detection using MTL

In this section, we will show studies on the use of MTL approaches in the field of anomaly detection, with the aim of improving the supervised processes.

Moon *et al.* [16] proposed a method for unsupervised anomaly detection in time-series sensor data of smart buildings. They used a model-agnostic MTL and the variational auto-encoder technique to adapt the model to a new target task with few unlabeled anomaly data. Entezami *et al.* [17] defined an unsupervised MTL method for health monitoring of civil structures for challenges such as large data with missing values and severe environmental changes. The MTL method is based on locally robust Mahalanobis-squared distances for online anomaly detection

Peng *et al.* [18] used an approach to spectrum anomaly detection in cognitive radio using MTL. The proposed method addresses the problem of the inability of existing deep learning-based spectral anomaly detection algorithms when are used directly across different frequency bands. The method involves using pre-training to analyze the differences between frequency bands, constructing an MTL dataset to find optimal model parameters, and fine-tuning the model using a small amount of target band data to detect anomalies. Tan *et al.* [19] proposed a self-supervised anomaly detection method that uses MTL to increase adaptability. The proposed method aims to improve sensitivity to subtle irregularities while maintaining robustness. The method is relevant for screening applications, and its effectiveness is demonstrated through experimental results.

Tavares *et al.* [20] proposed a method to extract meta-features from the event log and to recommend the most suitable encoding technique to improve anomaly detection performance. Their results showed that event log characteristics have different impacts on the representational capabilities. Dogo *et al.* [21] proposed a method for detecting anomalies in water quality, which is formulated as a classification problem in the presence of class imbalance. Sixteen single and static ensemble classification methods embedded with resampling strategies are optimized and compared, and six dynamic selection techniques are proposed and evaluated using an MTL approach.

With respect to the previous works, the objective of our work is to detect anomalies in the weight gained during cattle fattening, using an identification model based on an MTL architecture. MTL techniques were used to evaluate and select different ML algorithms to find the most suitable model.

2.2 Detection of anomalies in PLF

This section analyzes relevant research on the detection of abnormalities in the BP process, which are presented according to the object of study. The studies include the use of environmental and body temperature data, motion sensors, sound data, and activity monitoring, to detect anomalies in the health of cattle.

Animal behavior: Cai *et al.* [22] proposed a monitoring system for analyzing daily hog activity and abnormal behaviors in hog farms. The system uses a passive infrared detector and a high-accuracy acquisition system to collect data on daily hog activity, and uses an improved K-means clustering method to detect abnormal behavior during the night. The developed system provides data for the analysis and evaluation of the health, diseases, and environmental conditions of hog farms, which can affect fertility and productivity rates.

Animal welfare: Perrin *et al.* [23] evaluated an anomaly detection algorithm used in an automated surveillance system of cattle mortality. The method combined temporal regression and spatial cluster detection to identify clusters of spatial units showing an excess of deaths compared to their own historical fluctuations. The study simulated 1,000 outbreaks of a disease causing extra deaths in the French cattle population and applied the algorithm to identify clusters of spatial units showing an excess of deaths. The results indicated that the algorithm was able to identify unusual mortality clusters caused by an outbreak in certain conditions.

Kramer *et al.* [24] developed a fuzzy-logic model for the classification and control of lameness and mastitis, in cows, using data from the Futterkamp dairy-research farm of the Schleswig-Holstein Chamber of Agriculture. The fuzzy-logic model generated disease alerts using milk yield as the output variable, and as input data, dry-matter intake, dry-matter intake behavior (number of visits at the feeding trough, time spent at the feeding troughs), water intake, activity, and information about preliminary diseases. Sai *et al.* [25] developed an artificial intelligence module to estimate in a non-contact manner the body temperature of cattle, allowing for efficient individual monitoring of the health status of cattle. The module collected data on environment temperature, humidity, illuminance, and infrared images of cattle in a real-life environment.

Haladjian *et al.* [26] presented an approach to automatically detect cow lameness by monitoring changes in their gait, using a wearable motion sensor attached to their hind left leg. For that, the approach builds a model of a cow's usual walking pattern and detects deviations from this model. Results from a controlled experiment show that the approach can detect deviations in cows' gait with an accuracy of 91.1%. Wagner *et al.* [27] conducted a study on dairy cows, in which their activities were captured as time series by an indoor tracking system. The state of cows (diseases, stress, no problem) was manually labeled by animal caretakers, or by a sensor for ruminal pH (acidosis). Then their approach used a Fourier-based method to detect anomalies in time series.

Chung *et al.* [28] proposed a data-mining solution for the detection of oestrus, using sound data from Korean native cows (Bos taurus coreanea). They extracted the mel frequency cepstrum coefficients from sound data, with a feature-dimension reduction technique, and used a support vector machine for anomaly detection. The results indicated that this method can be used to detect estrus both economically (even with a cheap microphone) with an accuracy greater than 94%. Finally, García *et al.* [3] proposed an approach to detect anomalies in the cattle fattening processes. This approach used the historical record of animal weight to identify whether animals have gained the appropriate weight over time. They compared several ML techniques (Decision Tree, Gradient Boosting, K-Nearest Neighbors-based regression and Random Forest) in the task of anomalous weight detection, using Mean Absolute Error as quality metrics.

Pastures: Calera *et al.* [29] presented two approaches to monitor pasture quality based on multi-source info, being DEIMOS-1 the major satellite contributor. The first approach is based on in-depth monitoring of the crop phenology, characterized by means of the weekly Normalized Difference Vegetation Index. The second approach is focused on drought and other anomaly detection in crops and pastures. For pastures, a specific module has been designed to detect drought occurrence, by comparing current Normalized Difference Vegetation Index values with historical ones. Besides, they detected maximum livestock stocking rate, need of supplementary feeding, and overstocking risk.

Farm equipment: Park *et al.* [30] defined a mechanism to detect anomalies in pig house equipment using a recurrent neural network learning model, with data from sensors and environmental controllers. They predict malfunctions of each equipment, and when something goes wrong with the sensor, they use the difference between the predicted value and the measured value.

Early detection of abnormalities in animal husbandry is important for several reasons. First, it can improve production efficiency by allowing farmers to intervene quickly in situations that affect animal growth and health. Second, it can improve animal welfare by detecting and treating diseases and injuries before they become serious problems. In addition, early detection of abnormalities can help prevent the spread of disease to other animals on the farm, and in the wider community. Therefore, early detection of abnormalities is significant both for the health and welfare of the animals and for the economic sustainability of the livestock industry.

3 MTL Framework

MTL paradigm can be used to predict the performance of an ML algorithm on a specific task [31]. Thus, one goal of the MTL paradigm is to find the correlation between the characteristics of a dataset and the performance of different learning algorithms. With this information, a predictive meta-model can be defined to estimate the performance of different ML approaches on a dataset.

3.1 Conceptual architecture

The general architecture of the proposed framework is depicted in Figure 1. The association engine selects the best algorithm and parameterization by using a knowledge base. The knowledge base is a collection of meta-features describing datasets, models, and hyperparameters. Each time a new dataset is presented to the system, the association engine provides a suggestion of the most appropriate model (predictors, with their hyperparameters configuration). Immediately, the quality of the prediction is evaluated, and in case it is acceptable, then it is added to the knowledge base which is located in the Meta-Model. In this way, the identification system for anomaly detection is able to adapt to new data (few or many), using the MTL paradigm together with the metadata that is stored in the form of a meta-model base. In general, this identification system has been developed on the concept of MTL, so it consists of three main modules: learning, association, and adaptation.

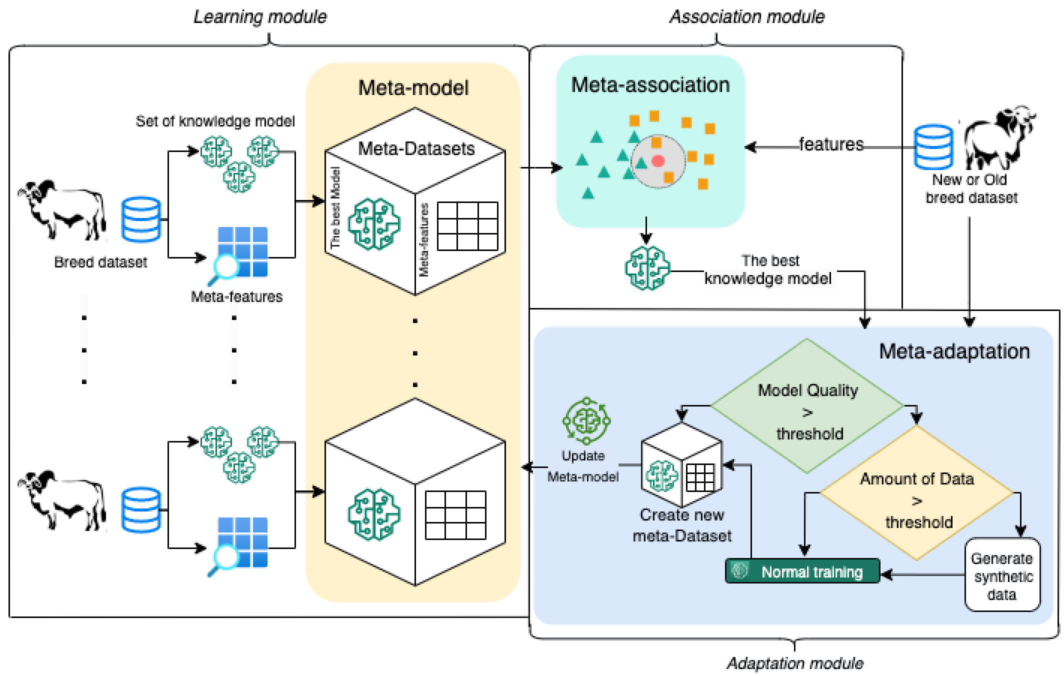


Figure 1: Meta Architecture for PLF

3.2 Learning module

It defines a knowledge base using metadata about previous learning tasks and learned models. It contains the characteristics of the datasets, e.g., mean, median, variance, and standard deviation of the variables, performance measures of each machine learning algorithm on those particular datasets, among others. Each metamodel represents the correspondence between the meta-characteristics describing the dataset, the predictive performance obtained by the group of learning algorithms when applied to these datasets, and the best ML model for that dataset. Thus, this module keeps mainly the results of different prediction algorithms on datasets, and meta-features of these datasets. In this context, the appropriate ML model for a new dataset can be selected using these metamodels.

Particularly, these metamodels are defined at the beginning based on the ML models to be adapted (predictive, diagnostic of prescriptive models) in the context where will be used the MTL framework. For the definition are considered the parameters of the techniques used, and the metrics to assess the quality of ML models. With this information, our MTL framework can evolve and improve its performance over time as more information is gathered.

3.3 Association module

The association module starts when a new dataset arrives for analysis (with little or a lot of information). At that moment, a set of meta-features describing the new dataset are extracted. Then, this module compares the features of the new dataset with the previous ones, using K-Nearest Neighbors (KNN) to place the new dataset in the closest cluster (for this, it is assumed that each metamodel in the knowledge base is a cluster). Now, it suggests the best model with its respective hyperparameters.

3.4 Adaptation module

In the adaptation module, the selected model is tested using the new data and its performance is evaluated. If the performance is not satisfactory (it does not pass a quality threshold), then the system evaluates whether the new dataset has enough data for a normal training. If it has too little data, then it proceeds to generate synthetic data and train. For both cases (enough data or synthetic data), then it ends up in normal training. Next, this module updates the meta-model (if any were selected), or builds and adds a new meta-model to the knowledge base (if it creates a new ML model). In general, this module acts when there is a new dataset (in this particular case, a new breed or crossbreed) or an old dataset that has time without being used. Then, it updates the meta-features and performance measures of current metamodels, or creates a new metamodel. In conclusion, this module creates/selects the best model and the best hyperparameters for a given dataset. In addition, the meta-model´s knowledge base is updated each time, allowing the system to continuously improve its performance over time.

4 Case Study

This section presents the experimental context of the case study and the instantiation of the MTL architecture for PLF in it.

4.1 Context

In cattle ranching, a system widely used by beef producers in the tropics is rotational grazing. Rotational grazing consists of dividing the entire surface area of a farm into more than two paddocks, while some remain occupied, the others are at rest [32]. This reduces the total grazing area, and forces cattle to consume forage uniformly, assigning batches of cattle to different paddocks [33] (see Figure 2).

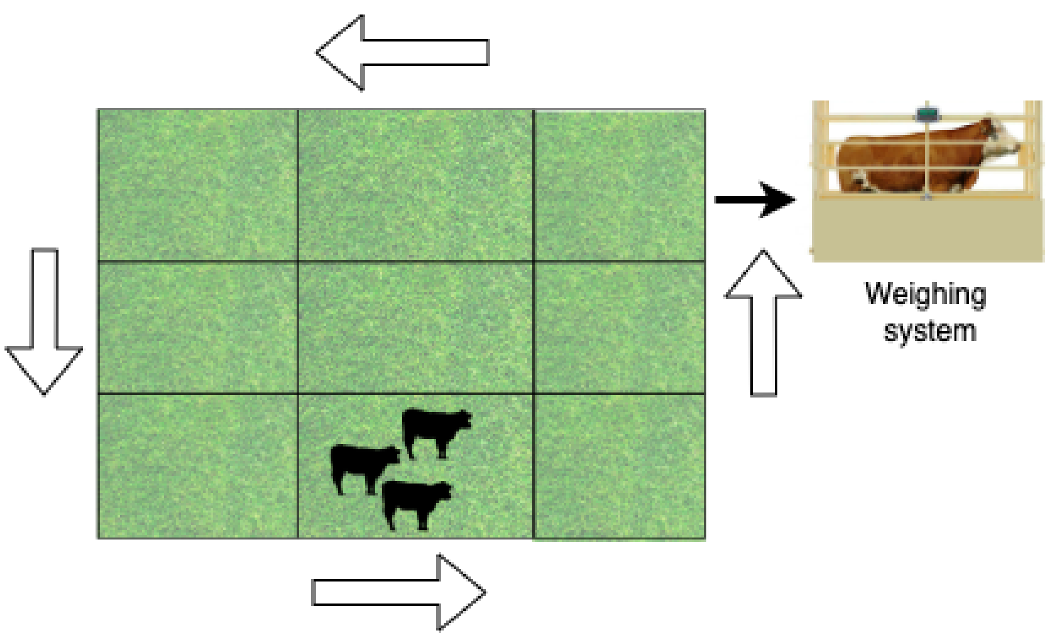


Figure 2: Rotational grazing system

Weighing cattle on rotational grazing is crucial for effective herd management because it is an important indicator of performance, and provides valuable information on grazing quality and other factors that can affect cattle growth and health. By monitoring cattle weight, producers can make informed decisions about grazing and feeding management, which can improve their farm's productivity and profitability.

In addition, cattle weights can also help detect health problems in cattle early, allowing producers to take preventative measures before they become serious problems. Regular monitoring of cattle weight can also help producers optimize grazing management, which can reduce the risk of overgrazing and improve long-term pasture and soil quality. In summary, weighing cattle on rotational grazing is an important practice that can improve cattle management and farm productivity.

In summary, it is important to weigh cattle on rotational grazing for several reasons:

- Monitoring cattle performance: Cattle weight is an important indicator of the performance of the BP, which can help producers make decisions about herd management, including the timing of the sale.
- Evaluating grazing success: Cattle weight can also provide information on grazing quality, feed availability, and other factors that may be affecting cattle growth.
- Early detection of health problems: Fluctuations in weight can be an early sign of health problems in cattle, allowing producers to act before they become major problems.
- Improve grazing: Weight information can also help producers make decisions about grazing management, including pasture rotation and the number of cattle in each section, which can improve cattle growth and health, as well as farm productivity.

In particular, we use real datasets in our case study, which are the records of animals raised and fattened at the "El Rosario" ranch, located in the municipality of Montería (Córdoba-Colombia).

4.2 Instantiation of the MTL architecture for PLF

This section describes the use of our MTL architecture in a monitoring system of the animal fattening process at the Rosario ranch for the detection of anomalies. In general, the system starts with the learning module where we have different animal breed datasets containing the variables gender, age, and weight. These breeds are crosses of the genetic groups Angus x Cebu (AC); Bon x Cebu (BC); Cebu x Angus x Cebu (CAC); Cebu x Cebu (CC); Holstein x Cebu (HC); Bon x Angus x Cebu (BAC), typical in Colombian farms. The data information for different animal breeds is described in Table 1, in which we see that their statistical metrics are quite close.

Table 1: Breed information

		Weight		
Breed	Count	**Mean**	**Min**	**Max**
AC	398	427.850369	165.219753	574.046996
BAC	398	410.001041	173.788039	541.459499
CAC	398	442.683480	181.033801	581.983351
BC	398	459.947979	156.834548	619.393828

The relationship between cattle weight and age data is crucial in detecting abnormalities and building effective models for livestock management, as can be seen in Figure 3.

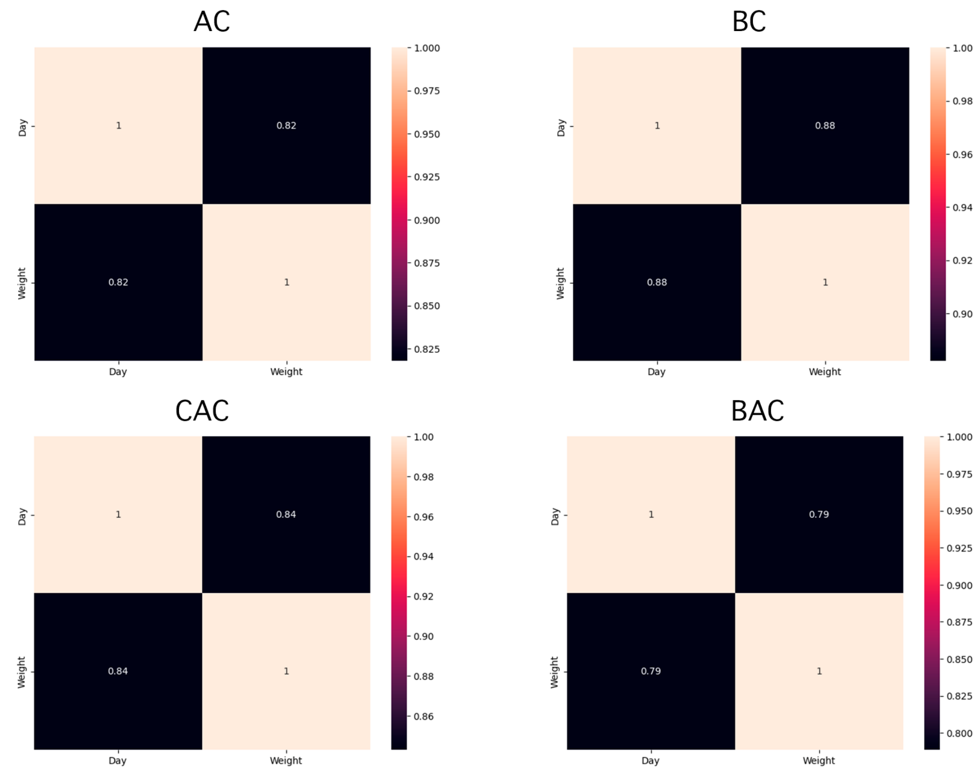


Figure 3: Correlation between age and weight

These data are highly correlated, and analyzing them together provides valuable information on the health status and performance of cattle. Therefore, cattle age is a key factor influencing their weight because as animals age, it is natural to expect their weight to increase. Thus, understanding how this relationship evolves at different stages of growth is essential.

The learning module refers to the process of acquiring knowledge and skills that make it possible to learn to learn. The goal of this module is to understand the patterns and strategies used to learn, and apply that knowledge to future learning situations. That involves selecting and preparing the data to learn, which includes, among other things, choosing and tuning the appropriate ML model for the specific problem being addressed.

For our particular case, for each breed, a set of knowledge models is trained using different ML techniques with the dataset features. From the set of models, the one with the best quality metrics is selected. With this model, the metamodel of this dataset is created, which will be stored in the knowledge base (see Table 2).

Three ML techniques were used in this first module: K-Neighbors Regressor (KNN), Gradient Boosting Regressor (GB), and Random Forest Regressor (RF). GridSearchCV is used as a hyperparameter optimization tool. This tool exhaustively searches for the optimal combination of hyperparameters of a model, to improve its performance and accuracy. Moreover, the features of the variables of the dataset are calculated. Finally, a meta-model is assembled for each dataset, which is composed of the best ML model and their meta-features (see Table 2).

Table 2: Meta-Model

knowledge			*Meta-features*							
Breed	Model	Best params	R^2	Median	Mean	Std	Var	Kurtosis	Entropy	Variation
AC	GB	learningrate:0.1 maxdepth:5 minsamplessp:5 nestimators:40	92	451.6	427.8	64.6	4175.25	1.20	10.62	0.15
BAC	RF	maxdepth:20 maxfeatures:3 maxleafnodes:25 nestimators:40	89	430.24	410.0	57.1	3262.80	1.49	10.62	0.13
CAC	GB	learningrate:0.1 maxdepth:5 minsamplessp:5 nestimators:40	93	470.04	442.6	72.5	5265.85	0.86	10.62	0.16
BC	GB	learningrate:0.1 maxdepth:5 minsamplessp:5 nestimators:40	95	492.2	459.9	87.5	7658.87	0.33	10.62	0.19

The following sections will explain the operation of the other modules of our architecture.

5 Experiments

Our architecture is divided into 3 modules, the first one was already instantiated in the previous section, and the next two are explained below. In this section, we will describe the behavior of the previously trained ML models in the MTL architecture, using data from the farm "Ganadería El Rosario". The duration of the experiments is a complete cycle of animal fattening on the farm before their sales.

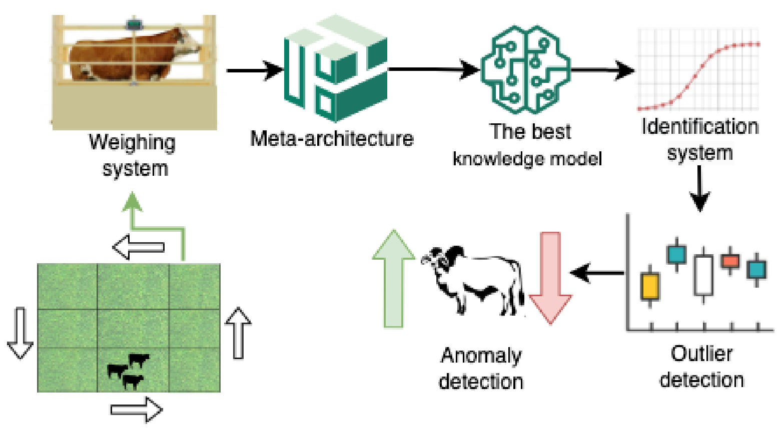


Figure 4: Schematic diagram of system behavior

The weighing process in rotational grazing involves the use of an animal weighing system (weigh scale) that is installed in a walkway or access corridor between the different grazing sections. The animals are guided to this weighing scale. During the process, the animals are weighed, and the data is recorded in a livestock management software.

This data is input to our architecture that will select the best model to build an ideal weight identification system. Subsequently, it will detect anomalies using outlier detection techniques (see Fig 4).

MTL tells ML how to learn, in order to learn from previous learning experiences rather than starting from scratch for each task. To show the versatility of our MTL architecture, we pose different scenarios that show how our proposal adapts to new and old datasets.

Thus, there are two main scenarios, the first is when the breed is known (old dataset) in the knowledge base and the second scenario is when it is an unknown breed (new dataset). For both scenarios, 2 sub-scenarios can be presented, the first would be when the model selected in the knowledge base has a good quality metric when using that dataset (it is not necessary to retrain), and the second when the metric quality is not good.

5.1 Scenario 1: Breed known by the knowledge base

5.1.1 Good quality of the model

If the breed is known (e.g., AC breed), then our MTL architecture searches in its knowledge base (Meta-model) the corresponding information. Subsequently, it selects the best model for that breed and makes the prediction. For this case, the best model is GB (see Table 3). Finally, it compares the ideal weight growth curve described by the AC breed cattle identification model with the current weights to detect anomalies (point cloud in Fig 5). In this scenario, the decision to use the GB model was based on its performance in quality metrics (see Table 4).

Table 3: Meta-model in the case of a breed known by the knowledge model

Breed	Model	Best params	R^2	Median	Mean	Std	Var	Kurtosis	Entropy	Variation
AC	GB	learningrate:0.1 maxdepth:5 minsamplessp:5 nestimators:40	92	451.6	427.8	64.6	4175.25	1.20	10.62	0.15

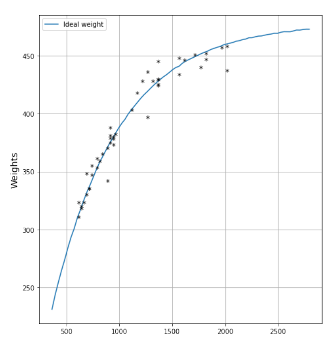


Figure 5: Ideal weight-curve versus current weight in AC breed

Since the curve represents the ideal temporal behavior of weight, points that deviate from it are potentially anomalous, but it is necessary to define the magnitude of the distance that separates an anomalous weight from one considered normal. To define this, multiple predictions of ideal weight are obtained for each combination of day, breed, and sex. Ideal weight intervals (not outliers) are then constructed to detect outliers using an isolation forest. This is an ML algorithm for anomaly detection that uses randomized decision trees to isolate anomalies in a dataset [34]. This allows for rapid detection of anomalous increases and decreases in animal weights at any desired time for decision-making.

Table 4: Decision metrics with good values

Breed	Model	R^2
AC	GB	92.5

5.1.2 Bad quality of the model

In the event that the quality of the metric is bad (see Table 5), a new training will be performed using new data (synthetic or this data) to adapt the selected model to these new situations described by the data. This means that the models can improve their predictive capability over time.

Table 5: Decision metrics with bad values

Breed	Model	R^2
AC	GB	57.2

Thus, the quality of the predictions is improved thanks to the meta-adaptation module. In this particular case, the new performance of the selected model after retraining is shown in Table 6.

Table 6: Improved decision metrics

Breed	Model	R^2
AC	GB	87.5

5.2 Scenario 2: Breed unknown by the knowledge base

In the case that a new breed arrives at the farm, or a new cross is created, the system will activate the second module (association) to select the best learned knowledge model for this specific dataset. This module extracts a set of meta-features from the dataset of the new breed to associate the model that best fits the specific features of the new data. Thus, this module automatically selects the best model based on the specific features of the dataset and its similarity with the features of the datasets with which previous models were trained.

In order to test the level of adaptation of our proposal, we removed the CAC, BAC and CC breeds from the knowledge base. Subsequently, the adaptation module is activated. Thus, first the quality of the model and the amount of data are evaluated. For this particular case, the model was not trained on the current BAC, CAC and CC breed, so the meta-model does not know these breeds. The association module is activated to search for the most similar group. Table 7 shows the input of the association module, and Table 8 shows to which group each breed is assigned.

Table 7: Feature inputs for the association module

Breed	Median	Mean	Std	Var	Kurtosis	Entropy	Variation
BAC	430.24	410.0	57.1	3262.80	1.49	10.62	0.13
CAC	470.04	442.6	72.5	5265.85	0.86	10.62	0.16
CC	471.631	455.190	45.666	2085.421	0.900	5.241	0.100

Table 8: Meta-Association

Breed	Median	Mean	Std	Var	Kurtosis	Entropy	Variation	Cluster
AC	451.6	427.8	64.6	4175.25	1.20	10.62	0.15	0
BAC	430.24	410.0	57.1	3262.80	1.49	10.62	0.13	0
CAC	470.04	442.6	72.5	5265.85	0.86	10.62	0.16	0
BC	492.2	459.9	87.5	7658.87	0.33	10.62	0.19	1
CC	471.6	455.1	45.6	2085.42	0.90	5.24	0.10	0

If the performance of the selected knowledge model is of bad quality for the new dataset, then this can have a negative impact on the decisions and results based on it. For this case, the MTL framework invoked by the fattening monitoring system will consider the following cases.

5.2.1 Good quality of the model

If the selected model has a quality higher than the quality threshold using the new dataset, then a new meta-model is created that contains this model as the best model and the meta-features of this new dataset. Subsequently, the knowledge base is updated to be aware of this new breed (see Table 9).

Table 9: New meta-model

Breed	Model	R^2	Median	Mean	Std	Var	Kurtosis	Entropy	Variation
AC	GB	92	451.6	427.8	64.6	4175.25	1.20	10.62	0.15
BC	GB	95	492.2	459.9	87.5	7658.87	0.33	10.62	0.19
BAC	GB	87	430.24	410.0	57.1	3262.80	1.49	10.62	0.13
CAC	GB	90	470.04	442.6	72.5	5265.85	0.86	10.62	0.16
CC	GB	90	471.6	455.1	45	2085.42	0.90	5.24	0.10

5.2.2 Bad quality of the model

For this case, if the performance of the selected model is of poor quality with the new data set, then two situations can occur:

- In the first situation, the amount of data is sufficient to train a model (see Table 10). Previously, we have removed the CC race from the knowledge dataset to test this situation). In this case, the system tests different ML techniques using a cross-validation process. Finally, a list of models with their quality metrics is obtained. The next step is to select the one with the highest performance (see Table 11) to update the meta-model with this new breed (see Table 12).

Table 10: Description of breed data CC

Breed	Median	Mean	Std	Var	Amount of data
CC	471.6	455.1	45.6	2085.42	976320

Table 11: List of models for the CC breed

Model	R^2	Select
RF	91.6	True
GB	89.2	False
DT	86.3	False
KNN	82.5	False

Table 12: New meta-model

Breed	Model	R^2	Median	Mean	Std	Var	Kurtosis	Entropy	Variation
AC	GB	92	451.6	427.8	64.6	4175.25	1.20	10.62	0.15
BC	GB	95	492.2	459.9	87.5	7658.87	0.33	10.62	0.19
BAC	GB	87	430.24	410.0	57.1	3262.80	1.49	10.62	0.13
CAC	GB	90	470.04	442.6	72.5	5265.85	0.86	10.62	0.16
CC	RF	91.6	471.6	455.1	45	2085.42	0.90	5.24	0.10

- In the second situation, the amount of data is not sufficient to train a model. Therefore, we proceed to generate synthetic data to be able to perform the training. Synthetic data has already been used in different areas [35]. Once the amount of data necessary for the traditional training is completed, then the steps described for the first situation are executed.

5.3 General Analysis

Some features of the MTL framework are the following. It can start from a set of models built with available historical cattle weight data, allowing the system to learn patterns and relationships between the data. On the other hand, it has the ability to adapt to face unforeseen situations (for example, new breeds) or the emergence of new variables, which gives it great flexibility and the ability to respond to different scenarios. This is especially important in the cattle industry, where changes in the environment and conditions can affect the weight of cattle and therefore their health.

In general, our approach can consider different scenarios in the adaptation process of ML models. When models fit the new data well, simply update your metamodels to that context, and when it doesn't, retrain the model with that new data (even if it's not enough to retrain, eventually generate synthetic data). As can be seen, all possible adaptation scenarios of ML models are considered. It can also be seen that our approach does not depend on specific ML techniques (random forest, regression, etc.). New models based on other techniques can be added. The same is true for ML model types. In this study, identification models were considered, but future work could consider diagnostic and optimization models, among others.

Finally, our proposal had a general average of R^2 of 90.8% for the different scenarios proposed. MTL enhances learning capabilities according to new data presented to it. In other words, our MLT framework is capable of learning to learn. In this context, our proposal has the capacity to continuously improve its anomaly detection capacity, even if the characteristics of the cattle or the environment in which they are found change over time.

6 Comparison with Other Works

In this section, we propose several criteria to compare our approach with similar works. Particularly, we define three criteria, which are:

- ***Criterion 1:*** The work includes adaptive modeling processes.
- ***Criterion 2:*** The work includes the transferability of the model to different domains or tasks.
- ***Criterion 3:*** The work includes incremental learning capabilities.

In Table 13, a qualitative comparison with related studies is made, based on previous criteria.

Table 13: Comparison with previous works.

	Criterion 1	***Criterion 2***	***Criterion 3***
[16]	✗	✓	✗
[17]	✗	✗	✗
[18]	✗	✗	✗
[19]	✗	✗	✗
[20]	✗	✓	✗
[21]	✗	✗	✗
This work	✓	✓	✓

For criterion 1, our work is the only one that meets. This criterion is important since the inclusion of adaptive ML-based modeling processes is a valuable feature in any work that involves machine learning. Adaptive modeling processes allow the model to change and improve over time by taking into account new data or changes in the underlying system being modeled. Otherwise, it would be static, that is, it is designed to remain fixed and unchanged once it is trained.

For criterion 2, the work [16] can be used in real-time facility management systems in smart buildings for several applications, including early detection of equipment failure, automated monitoring, and energy conservation. In addition, [20] proposes detecting anomalous instances in business processes to avoid resource waste and mitigate security issues. Finally, in our work, the transferability of the model is observed in the model's ability to adapt and generalize to new tasks. In our approach, the models can be retrained in new datasets (can represent specific tasks), to include the knowledge and experience inside of them to the original model. For criterion 3, our framework possesses incremental learning capability because our architecture can continuously learn when new datasets or information is received.

On the other hand, we have made a more specific comparison of works on the detection of anomalies in cattle weight. The paper of Segerkvist et al. [36] evaluated a method for monitoring the health of grazing cattle based on an unmanned, automatic precision weighing system that can be used on pasture. This system can generate alarms when animals show abnormal weight gain curves. The paper focused mainly on the detection of pasture-borne nematode parasite infections, which affect the weight gain of calves. Wagner et al. [37] used ML techniques to detect abnormal behaviour in cows with the Sub-Acute Ruminal Acidosis (SARA) disease, which is known to induce changes in behaviour. They used a positioning system to infer an animal's activity based on its position in relation to specific elements in the barn (alleys, feeder, and resting area), and defined ML models to predict activity on a given day. The work of Wagner et al. [27] detects the state of cows (diseases, estrus, no problem), which was trained using a dataset manually labeled by animal caretakers or by a sensor for ruminal pH (acidosis). They proposed a Fourier-based method to detect anomalies in this time series, which was compared with ML methods for time series classifications. As can be seen, the works are for specific diseases, or to determine behaviors, and do not propose a self-adaptive scheme of ML models, as is the basis of our proposal.

This work is relevant because it incorporates adaptive ML modeling processes, which allows the model to evolve and improve over time by including new data or changes in the underlying system. In addition, the model has the ability to transfer its knowledge to new tasks and can learn incrementally. These features make it useful for a variety of applications, including the detection of anomalies in cattle weights, which could have a significant impact on the cattle industry.

7 Conclusion and future work

While there are many applications of ML in livestock, the use of MTL in this industry is new and limited. In this paper, we have proposed an MTL architecture that can be used for PLF. Our architecture can use different ML models. The main advantage of our proposal is the ability to learn continuously as new data or information is received. In addition, it can update the existent models in its knowledge base, and its parameters, as new data is received. As well, our proposal can consider new datasets, maybe the reuse of trained models (it selects the most favorable one). Thus, one of the most outstanding qualities of our MTL architecture is its ability to adapt to new and old data over time.

Notably, our MTL architecture is the first to be used for cattle weight anomaly detection in the cattle industry. The metamodel knowledge base is continually updated to improve the identification/prediction models based on ML techniques over time. Our approach is effective in detecting livestock weight anomalies but can be adapted to different tasks.

However, our architecture also has some limitations that need to be addressed. First, it requires a large amount of data to train the initial models for the knowledge base, which may be difficult to obtain in some contexts. Second, it relies on the assumption that the ideal weight growth curve for a specific breed of cattle is similar in different contexts, which may not always be the case. Although the architecture we propose has some limitations, we believe that it can be used successfully in other scenarios of livestock management and animal productivity.

Future work is oriented to the development of specific autonomic cycles that allow self-monitoring and autonomous self-management processes in the context of PLF. In addition, another future work will be to test different synthetic data generation algorithms and evaluate their performance and behavior in our architecture.

Declarations

Data Availability Statement

Data will be made available in reasonable request

Funding and/or Conflicts of interests/Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Rodrigo García is supported by a PhD grant financed by Universidad EAFIT. All the authors would like to thank the "Vicerrectoría de Descubrimiento y Creación", of Universidad EAFIT, for their support on this research. Also, the authors want to thank the HAMADI 4.0 project, code 22-STIC-06, of the STIC-AmSud regional program, for the support of this work. Finally, the authors want to thank the student Manuela Zapata Mesa, from the Department of Mathematical Sciences, EAFIT University, for the scientific assistance provided during the realization of this work.

References

[1] N. C. Friggens, C. Disenhaus, and H. V. Petit, "Nutritional sub-fertility in the dairy cow: towards improved reproductive management through a better biological understanding," animal, vol. 4, no. 7, p. 1197–1213, 2010.

[2] X. Zhang, J. Ding, Y. Li, Q. Song, S. Li, M. Hayat, J. Zhang, and H. Wang, "The changes of inflammatory mediators and vasoactive substances in dairy cows' plasma with pasture-associated laminitis," BMC Veterinary Research, vol. 16, no. 1, 2020. cited By 6.

[3] R. García, J. Aguilar, M. Toro, and M. Jiménez, "Weight-identification model of cattle using machine-learning techniques for anomaly detection," in 2021 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 01–07, IEEE, 2021.

[4] D. Berckmans, "Precision livestock farming technologies for welfare management in intensive livestock systems," OIE Revue Scientifique et Technique, vol. 33, no. 1, p. 189 – 196, 2014. Cited by: 182.

[5] J. Aguilar and M. Cerrada, "Fuzzy classifier system and genetic programming on system identification problems," in Proceedings of the 3rd Annual Conference on Genetic and Evolutionary Computation, GECCO'01, (San Francisco, CA, USA), p. 1245–1251, Morgan Kaufmann Publishers Inc., 2001.

[6] L. Ljung, "Perspectives on system identification," Annual Reviews in Control, vol. 34, no. 1, pp. 1–12, 2010.

[7] R. García, J. Aguilar, M. Toro, A. Pinto, and P. Rodríguez, "A systematic literature review on the use of machine learning in precision livestock farming," Computers and Electronics in Agriculture, vol. 179, p. 105826, 2020.

[8] H. F. Gouda, F. A. M. Hassan, E. E. El-Araby, and S. A. Moawed, "Comparison of machine learning models for bluetongue risk prediction: a seroprevalence study on small ruminants," BMC Veterinary Research, vol. 18, no. 1, 2022. Cited by: 0; All Open Access, Gold Open Access, Green Open Access.

[9] C. Benitez, R. García, J. Aguilar, M. Jiménez, and H. Robles, "Supervision system of the fattening process of cattle in rotational grazing using fuzzy classification," in 2022 XVLIII Latin American Computer Conference (CLEI), pp. 1–10, 2022.

[10] J. X. Wang, "Meta-learning in natural and artificial intelligence," Current Opinion in Behavioral Sciences, vol. 38, pp. 90–95, 2021. Computational cognitive neuroscience.

[11] Y. Ma, S. Zhao, W. Wang, Y. Li, and I. King, "Multimodality in meta-learning: A comprehensive survey," Knowledge-Based Systems, vol. 250, p. 108976, 2022.

[12] F. Hutter, L. Kotthoff, and J. Vanschoren, Automated machine learning: methods, systems, challenges. Springer Nature, 2019.

[13] J. Aguilar, M. Jerez, E. Exposito, and T. Villemur, "CARMiCLOC: Context Awareness Middleware in Cloud Computing," in 2015 Latin American Computing Conference (CLEI), pp. 1–10, 2015.

[14] J. Aguilar, A. Garcès-Jimènez, N. Gallego-Salvador, J. A. G. De Mesa, J. M. Gomez-Pulido, and J. Garcìa-Tejedor, "Autonomic Management Architecture for Multi-HVAC Systems in Smart Buildings," IEEE Access, vol. 7, pp. 123402–123415, 2019.

[15] G. Morota, R. V. Ventura, F. F. Silva, M. Koyama, and S. C. Fernando, "Big data analytics and precision animal agriculture symposium: Machine learning and data mining advance predictive big data analysis in precision animal agriculture1," Journal of Animal Science, vol. 96, pp. 1540–1550, 03 2018.

[16] J. Moon, Y. Noh, S. Jung, J. Lee, and E. Hwang, "Anomaly detection using a model-agnostic meta-learning-based variational auto-encoder for facility management," Journal of Building Engineering, vol. 68, 2023. cited By 0.

[17] A. Entezami, H. Sarmadi, and B. Behkamal, "Long-term health monitoring of concrete and steel bridges under large and missing data by unsupervised meta learning," Engineering Structures, vol. 279, 2023. Cited by: 0.

[18] C. Peng, M. Zhang, W. Hu, and L. Wang, "A fast cross-band spectrum anomaly detection algorithm based on meta-learning," p. 420 – 424, 2022. Cited by: 2.

[19] J. Tan, T. Kart, B. Hou, J. Batten, and B. Kainz, "Metadetector: Detecting outliers by learning to learn from self-supervision," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 13166 LNCS, p. 119 – 126, 2022. Cited by: 0; All Open Access, Green Open Access.

[20] G. M. Tavares and S. B. Junior, "Process mining encoding via meta-learning for an enhanced anomaly detection," Communications in Computer and Information Science, vol. 1450 CCIS, p. 157 – 168, 2021. Cited by: 2.

[21] E. M. Dogo, N. I. Nwulu, B. Twala, and C. Aigbavboa, "Accessing imbalance learning using dynamic selection approach in water quality anomaly detection," Symmetry, vol. 13, no. 5, 2021. Cited by: 5; All Open Access, Gold Open Access.

[22] Y. Cai, L. Ma, and G. Liu, "A night-time anomaly detection system of hog activities based on passive infrared detector," Applied Engineering in Agriculture, vol. 35, no. 4, pp. 481–493, 2019.

[23] J.-B. Perrin, B. Durand, E. Gay, C. Ducrot, P. Hendrikx, D. Calavas, and V. Hénaux, "Simulation-based evaluation of the performances of an algorithm for detecting abnormal disease-related features in cattle mortality records," PLoS ONE, vol. 10, no. 11, 2015. Cited by: 5; All Open Access, Gold Open Access, Green Open Access.

[24] E. Kramer, D. Cavero, E. Stamer, and J. Krieter, "Mastitis and lameness detection in dairy cows by application of fuzzy logic," Livestock Science, vol. 125, no. 1, pp. 92–96, 2009. cited By 49.

[25] S. Ma, Q. Yao, T. Masuda, S. Higaki, K. Yoshioka, S. Arai, S. Takamatsu, and T. Itoh, "Development of an anomaly detection system for cattle using infrared image and machine learning," Sensors and Materials, vol. 32, no. 12, p. 4139 – 4149, 2020. Cited by: 3; All Open Access, Gold Open Access.

[26] J. Haladjian, Z. Hodaie, S. Nüske, and B. Brügge, "Gait anomaly detection in dairy cattle," vol. Part F132525, 2017. Cited by: 7.

[27] N. Wagner, V. Antoine, J. Koko, M.-M. Mialon, R. Lardy, and I. Veissier, "Comparison of machine learning methods to detect anomalies in the activity of dairy cows," in Foundations of Intelligent Systems (D. Helic, G. Leitner, M. Stettinger, A. Felfernig, and Z. W. Raś, eds.), (Cham), pp. 342–351, Springer International Publishing, 2020.

[28] Y. Chung, J. Lee, S. Oh, D. Park, H. Chang, and S. Kim, "Automatic detection of cow's oestrus in audio surveillance system," Asian-Australasian journal of animal sciences, vol. 26, no. 7, p. 1030–1037, 2013.

[29] A. Calera, F. Lozano, I. Campos, and A. Osann, "Selected farming services with satellite assistance: Mapping crops, optimizing water resources and assessing crop and pasture drought stress," pp. 62–72, 2013. cited By 0.

[30] H. Park, D.-H. Park, and S.-H. Kim, "Deep learning-based method for detecting anomalies of operating equipment dynamically in livestock farms," in 2020 International Conference on Information and Communication Technology Convergence (ICTC), pp. 1182–1185, 2020.

[31] M. Garouani, A. Ahmad, M. Bouneffa, M. Hamlich, G. Bourguin, and A. Lewandowski, "Using meta-learning for automated algorithms selection and configuration: an experimental framework for industrial big data," Journal of Big Data, vol. 9, p. 57, Apr 2022.

[32] D. Briske, J. Derner, J. Brown, S. Fuhlendorf, W. Teague, K. Havstad, R. Gillen, A. Ash, and W. Willms, "Rotational grazing on rangelands: Reconciliation of perception and experimental evidence," Rangeland Ecology Management, vol. 61, no. 1, pp. 3–17, 2008.

[33] E. J. Jacobo, A. M. Rodríguez, N. Bartoloni, and V. A. Deregibus, "Rotational grazing effects on rangeland vegetation at a farm scale," Rangeland Ecology Management, vol. 59, no. 3, pp. 249–257, 2006.

[34] H. Bansal, B. Chinagundi, P. S. Rana, and N. Kumar, "An ensemble machine learning technique for detection of abnormalities in knee movement sustainability," Sustainability, vol. 14, no. 20, 2022.

[35] R. D. Santos, J. Aguilar, and M. D. R-Moreno, "A synthetic data generator for smart grids based on the variational-autoencoder technique and linked data paradigm," in 2022 XVLIII Latin American Computer Conference (CLEI), pp. 1–7, 2022.

[36] K. A. Segerkvist, J. Höglund, H. Österlund, C. Wik, N. Högberg, and A. Hessle, "Automatic weighing as an animal health monitoring tool on pasture," Livestock Science, vol. 240, p. 104157, 2020.

[37] N. Wagner, V. Antoine, M.-M. Mialon, R. Lardy, M. Silberberg, J. Koko, and I. Veissier, "Machine learning to detect behavioural anomalies in dairy cows under subacute ruminal acidosis," Computers and Electronics in Agriculture, vol. 170, p. 105233, 2020.