Detection of carbapenem resistance in *Klebsiella pneumoniae* using convolutional vision transformers and MALDI-TOF proteomic profiles.

Valentina Salazar Marin

Advisor: Geysson Javier Fernández García
Co-advisor: Mario Alejandro Bravo Ortíz

Master Programme in Biosciences (Master Programme in Biosciences)
Res. 11910 of November 14, 2019. Valid until November 13, 2026. SNIES 108453

School of Applied Sciences and Engineering
(School of Applied Sciences and Engineering)
EAFIT University
October 2, 2025

Summary

Antimicrobial resistance is a growing global health problem, significantly increasing morbidity, mortality and healthcare costs. Traditionally, the identification of antibiotic resistance is based on phenotypic methods such as agar diffusion or automated systems such as VITEK, which require 24-72 hours to yield definitive results, delaying appropriate patient management. In this context, the need for faster and more accurate diagnostic aid strategies arises. This study explores the integration of artificial intelligence (AI) and mass spectrometry techniques for the classification of carbapenem-resistant *Klebsiella pneumoniae* strains. Specifically, matrix-assisted laser desorption/ionization-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) data were used to generate proteomic profiles potentially associated with resistance mechanisms. However, the complexity and high volume of these data make the use of AI tools capable of identifying robust patterns indispensable. A convolutional vision transformer (CVT) model was implemented to classify carbapenem-resistant *Klebsiella pneumoniae* strains from a set of 180 proteomic spectra collected by Synlab Colombia. The CVT model outperformed traditional convolutional neural networks and other automated learning approaches, achieving higher accuracy and stability. Grad-CAM visualization improved model interpretability by identifying key spectral regions associated with resistance. The results highlight the potential of Vision Transformers in microbiological diagnostics by significantly reducing resistance detection time and contributing to a timelier clinical response. Future studies should explore the applicability of this methodology on other resistant pathogens to improve global surveillance efforts against antimicrobial resistance.

TABLE OF CONTENTS

Chapter 1 5
Background 5
References 11
Chapter 2 14
1. Materials and Methods 14
1.1 Description of data 14
1.2 Database 14
2. Machine Learning Model Building 15
2.1 Data Processing for Machine Learning Models 15
2.2 Training and evaluation of Machine Learning models 16
2.2.1 Logistic regression (Logreg_cls) 17
2.2.2 Random Forest (RF) 17
2.2.3 K nearest neighbors (KNN) 17
2.2.4 Support Vector Machine (SVM) 18
2.2.5 Naive Bayes Gaussian (NBC) 18
2.2.6 Boosting gradient (Gradboost) 18
2.2.7 Light Gradient Boosting Machine (Lgbm) 18
2.2.8 Decision Trees (Dtree) 19
2.2.9 Multinomial Naive Bayes (NB) 19
2.2.10 XGBoost (Xgb) 19
2.2.11 Quadratic Discriminant Analysis (QDA) 19
2.2.12 AdaBoost Classifier (Adaboost) 20
2.2.13 Extra tree Classifier (Extratree) 20
3. Deep Learning model building 21
3.1 Model used 22
3.2 Parameters and Hyperparameters 30
3.2.1 Batch Normalization 30
3.2.2 Global Average Pooling 31
3.2.3. Softmax 31
3.2.4 Leaky ReLU 32
3.3 Transformer hyperparameters 33
3.4 Transfer Learning 34

3.4.1 VGG16 34
3.4.2 InceptionResNetV2 34
3.4.3 DenseNet201 35
3.4.4 EfficientNetB7 35
3.4.5. EfficientNetV2L 36
3.4.6 ConvNeXtXLarge 36
3.5 Fully Connected 36
3.6 Metrics 37
3.6.1 Accuracy 37
3.6.2 Precision 37
3.6.3 Sensitivity (Recall) 37
3.6.4 F1 Score 37
3.6.5 Support 38
3.6.6 Confusion Matrix 38
3.6.7 Cross-Validation 38
3.7 Hardware and resources 38
Results 39
1.Results and performance analysis of Machine Learning models 39
1.1 Evaluation through cross-validation 39
1.2 Evaluation by confusion matrix 40
2. Performance of the Convolutional Vision Transformer Model 42
Discussion 45
Conclusion 50
References 52
Acknowledgments 56

Chapter 1

Background

Antibiotic resistance is a growing global health crisis that significantly impacts morbidity and mortality rates associated with bacterial infections, as well as the economic burden of their treatment and control (Breijyeh et al., 2020). The World Health Organization (WHO) has recognized antimicrobial resistance as one of the top 10 global threats to public health (Widmer, 2022). Currently, resistant microorganisms are responsible for approximately 700,000 deaths per year worldwide, and projections indicate that this figure could exceed 10 million deaths per year by 2050 (Huemer et al., 2020).

Beyond its impact on mortality, antibiotic resistance imposes substantial economic costs. The Infectious Diseases Society of America (IDSA) estimates that treatment of resistant infections generates an annual expenditure of between $21 billion and $34 billion. A 2020 study found that hospital stays for patients with resistant bacterial sepsis involved an additional cost of $12,442 USD per patient compared to infections caused by susceptible bacteria (Aguilar et al., 2023).

The main microorganisms associated with this problem are *Staphylococcus aureus, Escherichia coli, Klebsiella pneumoniae, Pseudomonas aeruginosa and Acinetobacter baumannii*. These pathogens are responsible for more than 452,000 antibiotic resistance-related deaths annually in the Americas region (Aguilar et al., 2023). At the hospital level, carbapenem resistance in *K. pneumoniae* has increased exponentially, increasing mortality by 48% compared to infections caused by susceptible strains (Vera-Leiva et al., 2017). Moreover, infections caused by these resistant strains not only elevate mortality, but also significantly increases hospitalization costs and prolongs hospital stay by an average of 2 to 12.7 days (Allel et al., 2023).

Klebsiella pneumoniae has developed various mechanisms of antibiotic resistance, among which are alterations of porins, membrane proteins responsible for antibiotic transport, whose reduction or mutation limits the entry of these agents into the bacterial cell. In addition, the bacteria produce beta-lactamases, a group of enzymes capable of hydrolyzing β-lactam antibiotics such as penicillin, cephalosporins and carbapenemics. Within these, carbapenemases represent a subgroup with an extended spectrum of hydrolysis, which allows them to inactivate carbapenemics, one of the main therapeutic options against *K. pneumoniae* infections (Vera-Leiva et al., 2017) .

The increase in the production of carbapenemases has significantly complicated the treatment of these infections. According to the World Health Organization (WHO), by 2014, the rate of resistance to carbapenemics in strains of *K. pneumoniae* exceeded 50%, which has led to an increase in mortality and morbidity associated with infections caused by this bacterium (Vera-Leiva et al., 2017).

Laboratory diagnosis of resistance is traditionally carried out by phenotypic methods, such as broth microdilution and agar diffusion methods. Although these methods offer high accuracy, they require 24 to 72 hours to provide definitive results. This delay limits the ability of clinicians to initiate appropriate antibiotic treatment from the onset of infection, which can increase mortality and encourage the spread of resistant strains (Jeon et al., 2022). Several studies have evidenced that each hour of delay in the implementation of effective treatment in patients with systemic infection (sepsis), increase mortality by 7.6%, which underlines the importance of having rapid diagnostic techniques that allow early detection of resistance and optimize clinical management (Vila et al., 2017).

One of the most promising technologies for accelerating bacterial identification and resistance detection is matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). This technique allows rapid identification of microorganisms by analyzing their unique protein profiles, which significantly reduces diagnostic turnaround times compared to traditional methods (Gato et al., 2021).

The principle of MALDI-TOF MS is based on the generation of characteristic spectra within a range of 2,000 to 20,000 Da, which mainly represent ribosomal proteins and form a peptide fingerprint specific to each microorganism (Gato et al., 2023). This fingerprint is compared with reference databases, allowing the typing of isolates at the genus and species level. Due to its speed and reliability, MALDI-TOF MS has been incorporated as a routine tool in clinical microbiology, providing preliminary results that facilitate therapeutic decision-making.

Despite the success of MALDI-TOF MS in the identification of bacterial species, its application in the detection of antibiotic resistance profiles remains limited. However, recent studies have demonstrated its potential as a complementary tool for the discrimination of resistant strains, based on the identification of characteristic mass peaks in the spectra obtained.(Huang et al., 2020).

An example of this is the detection of methicillin-resistant *Staphylococcus aureus* (MRSA), where it has been reported that certain mass peaks allow differentiation between resistant and sensitive strains. In particular, a peak has been identified between m/z 2411 and 2415, corresponding to a small peptide called PSM-mec, highly specific for MRSA strains positive for mecA, the gene responsible for methicillin resistance. However, the presence of a single mass peak or the combination of several does not guarantee universal identification of resistance, as only 15.7% to 23.2% of MRSA strains present the peak between m/z 2411 and 2415, indicating that this approach may be limited to certain subgroups within the bacterial population (Kong et al., 2022).

The integration of artificial intelligence (AI) with MALDI-TOF MS has enhanced its analytical capabilities, allowing automation of the interpretation of complex data and improving the accuracy of pathogen identification. The increasing availability of large-scale microbiological datasets has further driven the application of AI in clinical microbiology, particularly in combination with technologies such as MALDI-TOF MS, which generates a vast volume of proteomic spectral data (Gato et al., 2021). AI-based models excel at recognizing statistical relationships within data, filtering out uninformative variables, and

identifying key spectral features, thus improving the accuracy of bacterial species classification and resistance detection (Weis et al., 2020).

Machine learning and deep learning algorithms have demonstrated exceptional performance in detecting complex patterns in clinical data, in particular genomic and proteomic information of microorganisms (Kim et al., 2022; Peiffer-Smadja et al., 2020). This approach has not only optimized the identification of bacterial species, but has also allowed exploring its potential in the detection of antimicrobial resistance, identifying spectral patterns associated with the production of resistant enzymes. An example of this is the analysis of *Klebsiella pneumoniae*, where it has been found that peaks in the m/z 11109 ratio correlate with the presence of a beta-lactamase variant, allowing the discrimination of some carbapenem-resistant *K. pneumoniae* (CRKP) strains. In this context, Huang et al. (2020) were able to classify 46 CRKP and 49 carbapenemicsensitive *K. pneumoniae* (CSKP) isolates with an accuracy of 97%, using machine learning models based on the spectra obtained.

Classical machine learning models, such as support vector machines (SVM) and random forests (RF), have demonstrated remarkable performance when applied to spectral data obtained by MALDI-TOF MS. In a study by Gato et al., (2023), an accuracy of 97.83%, with a sensitivity of 100%, a specificity of 96.73% and an F1 score of 96.85% was reported in the detection of carbapenemase-producing *K. pneumoniae* strains. Similarly, the combination of MALDI-TOF MS with the LightGBM algorithm has allowed the identification of *K. pneumoniae* with an accuracy of 83.61 % and an area under the curve (AUC) of 0.84 (Yu et al., 2023). Other works such as Zhang et al., (2023) and Zeng et al., (2023) have shown comparable performances using deep neural networks and machine learning models. In the study by Zhang et al. an optimized artificial neural network (ANN) was used to classify resistant strains of *K. pneumoniae* with 84% accuracy, while Zeng et al. combined MALDI-TOF MS with a LASSO logistic regression model, managing to identify key discriminant features within the spectra with 87% accuracy in Imipenem resistance classification.

The use of genomic data as a tool for antimicrobial resistance prediction has also been explored. In this context, Condorelli et al. (2024) employed machine learning models, such

as the k-nearest neighbor algorithm, achieving accuracies above 90% in predicting resistance to 10 different antibiotics from whole genomic sequences of *K. pneumoniae*. Complementarily, deep learning has demonstrated comparable performance, as evidenced by the study of López-Cortés et al. (2024), who developed a deep neural network called MSDeepAMR. This model was trained directly on raw mass spectra, focusing on common clinical pathogens such as *Escherichia coli*, *Klebsiella pneumoniae* and *Staphylococcus aureus*, under different resistance profiles. MSDeepAMR showed good performance in the classification of resistant strains, reaching areas under the ROC curve above 0.83 in most of the cases analyzed.

The use of machine and deep learning models for bacterial resistance classification is a powerful tool for optimizing diagnosis and clinical decision making. However, it is critical to consider the characteristics of the data generated by MALDI-TOF MS. These data consist of hundreds or thousands of mass-to-charges (m/z) ratios per sample, each with its respective intensity level, which generates a high dimensionality problem, where the number of variables is considerably larger than the sample size (Huang et al., 2020). As a consequence, many pattern classification algorithms may fail, as the differences between spectral peaks are minimal and do not always reflect biological variability, but rather technical variability. Therefore, it is essential to apply adequate preprocessing to spectra before implementing artificial intelligence models, a challenge that we directly address in the present study.

This study worked with data obtained from the Synlab laboratory, where the identification of microorganisms is routinely performed, but without the use of advanced spectral processing techniques. To overcome the classification challenges derived from the high dimensionality of the m/z and intensity data, this study adopted a two-fold strategy. In a first phase, various Machine Learning models were applied directly on the numerical matrix, which revealed significant limitations in classification accuracy. In order to overcome this limitation and compensate for the variability present in the data, an innovative approach was adopted: the use of the spectra images in conjunction with the Convolutional Vision Transformer (CVT) model. This methodology offers a more robust analysis, being able to

efficiently capture both local and global proteomic features, which significantly optimizes the differentiation between sensitive and resistant strains.

The integration of MALDI-TOF mass spectrometry with artificial intelligence represents a significant advance in the detection of antimicrobial resistance. This combination provides a classification tool that optimizes diagnosis, making it faster and more accurate, which in turn improves clinical decision making. Future efforts should focus on implementing these technologies in real-world clinical settings and developing models that consider the heterogeneity of microbiological and clinical data. This will enable. improved predictive performance and adaptability of systems in different healthcare settings.

References

Aguilar, G. R., Swetschinski, L. R., Weaver, N. D., Ikuta, K. S., Mestrovic, T., Gray, A. P., Chung, E., Wool, E. E., Han, C., Hayoon, A. G., Araki, D. T., Abdollahi, A., Abu-Zaid, A., Adnan, M., Agarwal, R., Dehkordi, J. A., Aravkin, A. Y., Areda, D., Azzam, A. Y., ... Naghavi, M. (2023). The burden of antimicrobial resistance in the Americas in 2019: a cross-country systematic analysis. *The Lancet Regional Health - Americas*, *25.* https://doi.org/10.1016/j.lana.2023.100561. https://doi.org/10.1016/j.lana.2023.100561.

Allel, K., Stone, J., Undurraga, E. A., Day, L., Moore, C. E., Lin, L., Furuya-Kanamori, L., & Yakob, L. (2023). The impact of inpatient bloodstream infections caused by antibiotic-resistant bacteria in low- and middle-income countries: A systematic review and meta-analysis. *PLOS Medicine*, *20*(6), e1004199. https://doi.org/10.1371/JOURNAL.PMED.1004199. https://doi.org/10.1371/JOURNAL.PMED.1004199

Breijyeh, Z., Jubeh, B., & Karaman, R. (2020). Resistance of Gram-Negative Bacteria to Current Antibacterial Agents and Approaches to Resolve It. *Molecules*, *25*(6). https://doi.org/10.3390/MOLECULES25061340. https://doi.org/10.3390/MOLECULES25061340

Condorelli, C., Nicitra, E., Musso, N., Bongiorno, D., Stefani, S., Gambuzza, L. V., Carchiolo, V., & Frasca, M. (2024). Prediction of antimicrobial resistance of Klebsiella pneumoniae from genomic data through machine learning. *PLOS ONE*, *19*(9), e0309333. https://doi.org/10.1371/JOURNAL.PONE.0309333.

Gato, E., Arroyo, M. J., Méndez, G., Candela, A., Rodiño-Janeiro, B. K., Fernández, J., Rodríguez-Sánchez, B., Mancera, L., Arca-Suárez, J., Beceiro, A., Bou, G., & Oviaño, M. (2023). Direct Detection of Carbapenemase-Producing Klebsiella pneumoniae by MALDI-TOF Analysis of Full Spectra Applying Machine Learning. *Journal of Clinical Microbiology*, *61*(6), e0175122. https://doi.org/10.1128/JCM.01751-22. https://doi.org/10.1128/JCM.01751-22

Gato, E., Constanso, I. P., Candela, A., Galán, F., Rodiño-Janeiro, B. K., Arroyo, M. J., Méndez, G., Mancera, L., Alioto, T., Gut, M., Gut, I., Álvarez-Tejado, M., Rodríguez-Sánchez, B., Bou, G., & Oviaño, M. (2021). An Improved Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry Data Analysis Pipeline for the Identification of Carbapenemase-Producing Klebsiella pneumoniae. *Journal of Clinical Microbiology*, *59*(7). https://doi.org/10.1128/JCM.00800-21/SUPPL_FILE/JCM.00800-21-S0001.PDF

Huang, T. S. S., Lee, S. S. J., Lee, C. C., & Chang, F. C. (2020). Detection of carbapenem-resistant Klebsiella pneumoniae on the basis of matrix-assisted laser desorption ionization time-of-flight mass spectrometry by using supervised machine learning

approach. *PloS One*, *15*(2). https://doi.org/10.1371/JOURNAL.PONE.0228459. https://doi.org/10.1371/JOURNAL.PONE.0228459

Huemer, M., Shambat, S. M., Brugger, S. D., & Zinkernagel, A. S. (2020). Antibiotic resistance and persistence-Implications for human health and treatment perspectives. *EMBO Reports*, *21*(12). https://doi.org/10.15252/EMBR.202051034. https://doi.org/10.15252/EMBR.202051034

Jeon, K., Kim, J. M., Rho, K., Jung, S. H., Park, H. S., & Kim, J. S. (2022). Performance of a Machine Learning-Based Methicillin Resistance of Staphylococcus aureus Identification System Using MALDI-TOF MS and Comparison of the Accuracy according to SCCmec Types. *Microorganisms 2022, Vol. 10, Page 1903*, *10*(10), 1903. https://doi.org/10.3390/MICROORGANISMS10101903. https://doi.org/10.3390/MICROORGANISMS10101903.

Kim, J. I., Maguire, F., Tsang, K. K., Gouliouris, T., Peacock, S. J., McAllister, T. A., McArthur, A. G., & Beiko, R. G. (2022). Machine Learning for Antimicrobial Resistance Prediction: Current Practice, Limitations, and Clinical Perspective. *Clinical Microbiology Reviews*, *35*(3). https://doi.org/10.1128/CMR.00179-21. https://doi.org/10.1128/CMR.00179-21

Kong, P. H., Chiang, C. H., Lin, T. C., Kuo, S. C., Li, C. F., Hsiung, C. A., Shiue, Y. L., Chiou, H. Y., Wu, L. C., & Tsou, H. H. (2022). Discrimination of Methicillin-Resistant Staphylococcus aureus by MALDI-TOF Mass Spectrometry with Machine Learning Techniques in Patients with Staphylococcus aureus Bacteremia. *Pathogens*, *11*(5). https://doi.org/10.3390/PATHOGENS11050586/S1

López-Cortés, X. A., Manríquez-Troncoso, J. M., Hernández-García, R., & Peralta, D. (2024). MSDeepAMR: antimicrobial resistance prediction based on deep neural networks and transfer learning. *Frontiers in Microbiology*, *15.* https://doi.org/10.3389/FMICB.2024.1361795/PDF. https://doi.org/10.3389/FMICB.2024.1361795/PDF

Peiffer-Smadja, N., Rawson, T. M., Ahmad, R., Buchard, A., Pantelis, G., Lescure, F. X., Birgand, G., & Holmes, A. H. (2020). Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. *Clinical Microbiology and Infection*, *26*(5), 584-595. https://doi.org/10.1016/J.CMI.2019.09.009/ATTACHMENT/4EA59A1A-A7FB-4287-BD7C-AB65F1E449C7/MMC1.DOCX.

Vera-Leiva, A., Barría-Loaiza, C., Carrasco-Anabalón, S., Lima, C., Aguayo-Reyes, A., Domínguez, M., Bello-Toledo, H., González-Rocha, G., Vera-Leiva, A., Barría-Loaiza, C., Carrasco-Anabalón, S., Lima, C., Aguayo-Reyes, A., Domínguez, M., Bello-Toledo, H., & González-Rocha, G. (2017). KPC: Klebsiella pneumoniae carbapenemase, main carbapenemase in enterobacteria. *Revista Chilena de Infectología*, *34*(5), 476-484. https://doi.org/10.4067/S0716-10182017000500476.

Vila, J., Gómez, M. D., Salavert, M., & Bosch, J. (2017). Rapid diagnostic methods in clinical microbiology: clinical needs. *Infectious Diseases and Clinical Microbiology*, *35*(1), 41-46. https://doi.org/10.1016/J.EIMC.2016.11.004.

Weis, C. V., Jutzeler, C. R., & Borgwardt, K. (2020). Machine learning for microbial identification and antimicrobial susceptibility testing on MALDI-TOF mass spectra: a systematic review. *Clinical Microbiology and Infection*, *26*(10), 1310-1317. https://doi.org/10.1016/J.CMI.2020.03.014.

Widmer, A. F. (2022). *Emerging antibiotic resistance: Why we need new antibiotics!* https://doi.org/10.57187/smw.2022.40032.

Yu, J., Lin, Y. T., Chen, W. C., Tseng, K. H., Lin, H. H., Tien, N., Cho, C. F., Huang, J. Y., Liang, S. J., Ho, L. C., Hsieh, Y. W., Hsu, K. C., Ho, M. W., Hsueh, P. R., & Cho, D. Y. (2023). Direct prediction of carbapenem-resistant, carbapenemase-producing, and colistin-resistant Klebsiella pneumoniae isolates from routine MALDI-TOF mass spectra using machine learning and outcome evaluation. *International Journal of Antimicrobial Agents*, *61*(6), 106799. https://doi.org/10.1016/J.IJANTIMICAG.2023.106799. https://doi.org/10.1016/J.IJANTIMICAG.2023.106799

Zeng, Y., Wang, C., Ye, Q., Liu, G., Zhang, L., Wan, J., & Zhu, Y. (2023a). Machine learning model of imipenem-resistant Klebsiella pneumoniae based on MALDI-TOF-MS platform: An observational study. *Health Science Reports*, *6*(9), e1108. https://doi.org/10.1002/HSR2.1108.

Zhang, Y. M., Tsao, M. F., Chang, C. Y., Lin, K. T., Keller, J. J., & Lin, H. C. (2023a). Rapid identification of carbapenem-resistant Klebsiella pneumoniae based on matrix-assisted laser desorption ionization time-of-flight mass spectrometry and an artificial neural network model. *Journal of Biomedical Science*, *30*(1), 1-10. https://doi.org/10.1186/S12929-023-00918-2/FIGURES/5. https://doi.org/10.1186/S12929-023-00918-2/FIGURES/5.

Chapter 2

1. Materials and Methods

1.1 Description of data

1.2 Database

This retrospective study was conducted using data previously collected by Synlab Colombia. The dataset was obtained by searching WHONET, a free software developed by the WHO Collaborating Centre for Antimicrobial Resistance Surveillance. WHONET collects data from microbiology laboratories and is specifically designed to analyze antibiotic susceptibility and resistance test results. As part of a national and international surveillance network, it facilitates antimicrobial resistance monitoring, which helps to understand local microbial epidemiology, detect hospital outbreaks, and guide antimicrobial selection.

Clinical isolates of *Klebsiella pneumoniae* registered between 2022 and 2023 were selected from WHONET. The antibiotic susceptibility profile of each isolate was analyzed to classify it as susceptible or resistant to carbapenemics. Mass spectra corresponding to each isolate were subsequently exported from the MALDI-TOF Biotyper (Bruker), which uses matrix-assisted laser desorption/ionization-assisted time-of-flight mass spectrometry with laser desorption/ionization (MALDI-TOF MS) to identify microorganisms by analyzing their unique ribosomal protein profiles.

The data export process was performed using flexAnalysis, a specialized software developed by Bruker for the management and visualization of spectra obtained from Flex series mass spectrometers. Proteomic spectra were extracted, including from carbapenemicsensitive and resistant isolates.

The spectral variables analyzed in this study include mass-to-charge ratio (m/z) and intensity. In the mass spectrum, the ***x-axis*** represents the m/z values, which correspond to the ratio of the mass of an ion to its electric charge. These values provide crucial information about the nature of the ions generated from the sample. On the other hand, the ***y-axis*** represents the intensity, indicating the relative abundance of ions detected for each m/z value, reflecting the frequency of the presence of a particular ion in the sample.

Each peak in the mass spectrum corresponds to a specific ion. Its position on the ***x-axis*** reveals the mass-to-charge ratio, and its height on the ***y-axis*** indicates its abundance. This combined information forms a detailed proteomic profile, essential for identifying and quantifying the biomolecules present in the sample, as illustrated in **Figure 1.**

2. Machine Learning Model Building

2.1 Data Processing for Machine Learning Models

A total of 180 spectra were exported for the initial analysis, of which 80 corresponded to resistant bacteria and 100 to sensitive bacteria. Considering only the m/z values and intensity of the spectra.

In order to prepare the data for analysis using Machine Learning models, a feature transformation and selection process was performed. First, a logarithmic reduction was applied to the m/z and intensity values in order to normalize the distribution of the data and reduce possible biases in the scale of the variables. Subsequently, an evaluation of the relevance of spectral peaks was carried out using two criteria:

- Student's t-test, used to identify statistically significant differences (*p-value* <0.05) between the resistant and sensitive spectra groups. Normality and homogeneity of variances were evaluated to ensure the validity of the statistical analyses and the proper interpretation of the results.

- Fold Change, was used to quantify the magnitude of the variation in peak intensity between the two groups, considering as significant a change of at least twofold, either an increase or a decrease.

Only those characteristics that presented significant differences in both tests were retained, ensuring that the data used for training the models were composed of the most relevant variables for the classification of bacterial resistance.

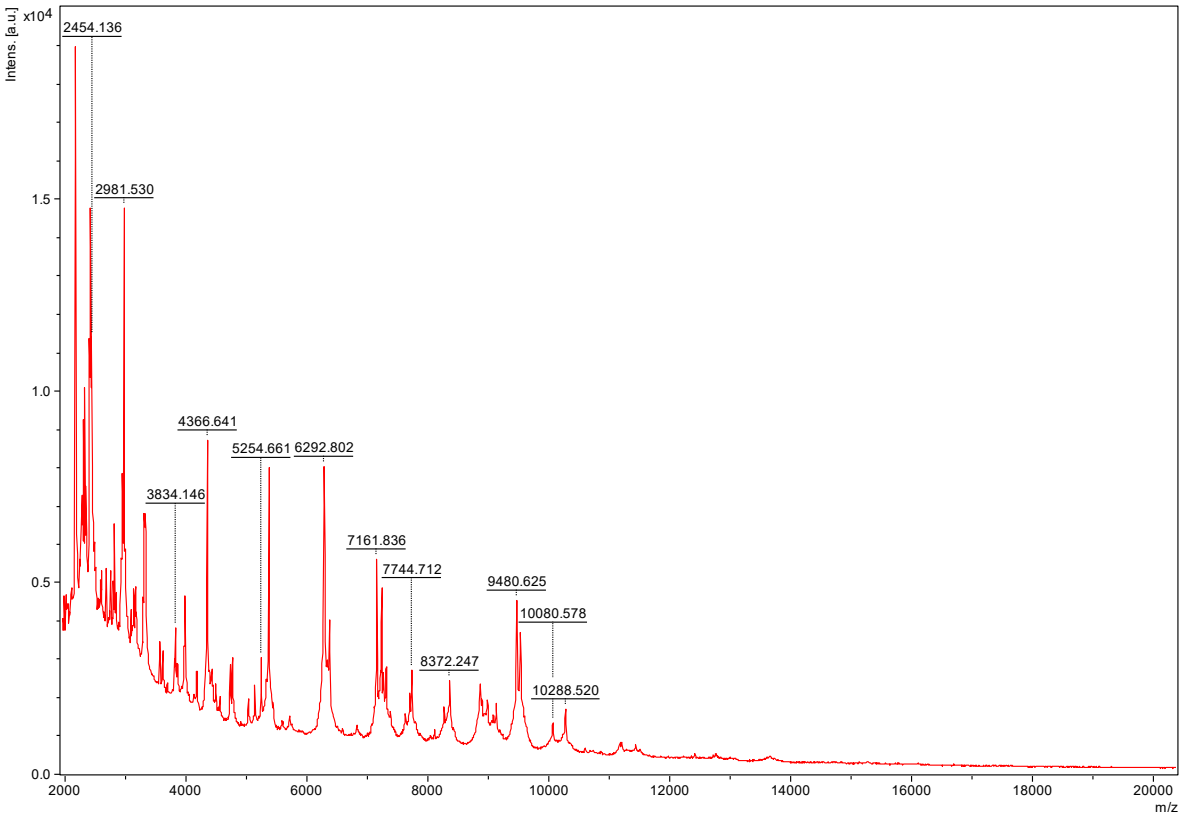

Figure 1. Visualization of the proteomic spectrum; x-axis mass/charge value, y-axis intensity value.

2.2 Training and evaluation of Machine Learning models

Once the data was preprocessed, the *Machine Learning* models were built. The data set was divided into two subsets: 80% was used for training and 20% was reserved exclusively for the final evaluation. As input variables, the transformed m/z and intensity values selected according to their statistical relevance were used.

During the training phase, a k=10 stratified cross validation (*K-fold*) was implemented within the training set, in order to estimate the performance of each model and prevent overfitting. Several supervised classification algorithms were trained, including: logistic regression, Random Forest, K-nearest neighbors (KNN), support vector machines (SVM), Naive Bayes (Gaussian and multinomial), decision trees, Gradient Boosting, LightGBM, XGBoost, AdaBoost, Extra Trees and quadratic discriminant analysis (QDA).

2.2.1 Logistic regression (Logreg_cls)

A supervised algorithm used for binary (or ordinal/multinomial multiclass) classification problems, which estimates the probability of class membership using the sigmoid function. It is used for binary classification, where the output can be one of two possible categories: Yes/No, True/False or 0/1. Unlike linear regression, its output is continuous, ideal for discrete results (GeeksforGeeks, 2025j).

2.2.2 Random Forest (RF)

Random Forest is an ensemble algorithm that combines multiple decision trees trained with random subsets of data and features to improve the accuracy and stability of predictions. It is applicable to both classification and regression tasks. In the case of classification, each tree makes an individual prediction and the final model assigns the class that receives the most votes. This technique reduces the risk of overfitting and improves generalization ability compared to a single decision tree(GeeksforGeeks, 2025g).

2.2.3 K nearest neighbors (KNN)

K-NN is a supervised nonparametric algorithm that classifies a new data based on the majority of classes of its *k* nearest neighbors in the feature space. It makes no assumptions about the distribution of the data and is especially useful in problems where the distance (Euclidean, Manhattan, or minkowsi) between observations is significant, it is used in classification and regression problems.(GeeksforGeeks, 2025c).

2.2.4 Support Vector Machine (SVM)

SVM is a supervised algorithm that searches for the optimal hyperplane that maximizes the margin between different classes. The larger the margin, the better the performance of the model with new and unpublished data. It can solve linear and nonlinear classification problems using *kernel* functions, which makes it powerful for complex and high-dimensional data.(GeeksforGeeks, 2025e)

2.2.5 Naive Bayes Gaussian (NBC)

It is a version of the Naive Bayes classifier that assumes that the numerical features follow a normal (Gaussian) distribution. It is based on Bayes' theorem with the strong (naive) assumption that no feature in the data set is related to the others. The algorithm calculates the variance and mean of each feature for each class during training. During the prediction stage, it determines to which class an instance is most likely to belong by calculating the probability of each of the evaluated classes.(GeeksforGeeks, 2025h).

2.2.6 Boosting gradient (Gradboost)

It is a boosting algorithm that combines different weak learners to create a robust predictive model. It is given by sequential training, usually decision trees, where each new model attempts to correct the errors of the previous one by optimizing a loss function. At each iteration, the algorithm calculates the gradient of the loss function with respect to the predictions and then trains a new weak model to minimize it. The predictions of the new model are added to the ensemble (the prediction of all models) and the process is repeated until a stopping criterion is satisfied.(GeeksforGeeks, 2025i).

2.2.7 Light Gradient Boosting Machine (Lgbm)

It uses multiple decision trees to make predictions. It starts with data organized by rows (instances) and columns (features). It then creates an initial model with a simple prediction, which usually has errors. From those errors, it calculates how to improve the model using a loss function and its derivatives (gradients); it builds decision trees based on the number of

samples in each leaf (instead of level by level), choosing splits that minimize error. This makes it possible to create deeper and more effective trees.
In addition, it includes techniques such as regularization and early stopping to avoid overfitting. Its high efficiency comes from using histogram data structures and optimizing memory usage. In the end, it combines all the trees to generate a more accurate prediction.(GeeksforGeeks, 2025b)

2.2.8 Decision Trees (Dtree)

These are supervised models that use a hierarchical structure of nodes to divide the data set into homogeneous subsets. Each node represents a decision based on a feature, and the leaves represent final predictions. They are suitable for both classification and regression tasks (GeeksforGeeks, 2025a).

2.2.9 Multinomial Naive Bayes (NB)

Designed for discrete features, this model is especially useful in text classification tasks. It assumes that features follow a multinomial distribution and are conditionally independent of each other. This means that the presence of one word does not affect the presence of another, which facilitates the use of the model.(GeeksforGeeks, 2025f)

2.2.10 XGBoost (Xgb)

It is an ensemble learning method, which combines multiple weak models to build a more robust one. It uses decision trees as a basis and incorporates them sequentially, where each new tree seeks to correct the mistakes made by the previous ones. This approach is known as *boosting*. One of its main advantages is that it allows fast training of models through parallel processing, which makes it ideal for working with large volumes of data. In addition, it offers extensive customization options, allowing to adjust its parameters to better adapt to the specific characteristics of each problem.(GeeksforGeeks, 2024)

2.2.11 Quadratic Discriminant Analysis (QDA)

QDA is a classifier that models each class with its own multivariate normal distribution, allowing it to capture quadratic decision boundaries. It is useful when classes have different variances and covariances, it does not assume that the correlation matrices of each class are equal, which allows it to construct more flexible decision boundaries, to describe each class with its own correlation matrix.(GeeksforGeeks, 2024).

2.2.12 AdaBoost Classifier (Adaboost)

AdaBoost or Adaptive Boosting is an ensemble algorithm that builds a strong model from several weak models, such as small decision trees; it starts by assigning equal weights to all samples in the training set and, at each iteration, adjusts those weights to give more importance to data that were misclassified. Each new model is trained by paying more attention to those errors, with the goal of correcting the failures of the previous model. In the end, all the individual models are combined to form a single, more accurate prediction. AdaBoost is effective at reducing bias and variance in classification tasks, although it can be sensitive to noisy data or data with outliers.(GeeksforGeeks, 2025d).

2.2.13 Extra tree Classifier (Extratree)

is an ensemble method similar to Random Forest, which combines multiple decision trees to improve model accuracy; its main difference is that it introduces more randomness in selecting the cut points during tree construction. Each tree is trained on the full data set, and at each node a subset of features is randomly selected. From these, the splitting threshold is randomly selected, rather than searching for the optimum. This strategy generates trees that are less correlated with each other, which improves generalization.(GeeksforGeeks, 2023).

For each model, performance metrics such as *accuracy*, *precision*, *recall* and F1 score, calculated from the average of the 10 cross-validation partitions, were evaluated. Subsequently, the best performing model was evaluated using the test set (remaining 20%) to obtain an independent and final estimate of its generalizability. In addition, a confusion

matrix was generated to analyze in detail the performance of the classifier in discriminating between sensitive and resistant strains.

3. Deep Learning model building

A total of 180 proteomic spectra images were extracted, including 100 from carbapenemicsensitive isolates and 80 from resistant isolates. Maintaining a balanced data set was key to ensuring the effectiveness of the artificial intelligence (AI) algorithms in this study**.**

A balanced data set is crucial for binary classification tasks such as carbapenem resistance detection. If the dataset were unbalanced, the model could be biased towards the majority class, reducing its ability to correctly identify resistant strains. Ensuring balance improves key evaluation metrics such as sensitivity, specificity and F1 score, providing a more reliable assessment of model performance.

In addition, this approach improves the model's ability to generalize new data, an essential aspect in clinical applications where failure to detect resistant strains could have serious consequences for patient health. By optimizing the training process, the model can effectively learn the distinguishing features of both classes, which increases its reliability in detecting bacterial resistance.

The spectral variables used for this model are those captured in the spectral image, m/z and intensity. With these two-dimensional matrices, we proceeded to use a Vision Transformer architecture. This architecture, originally designed for image classification, has been adapted in this study for the analysis of proteomic spectra. Unlike convolutional neural networks (CNNs), which extract local features by convolution operations, Vision Transformer divides proteomic spectra into small sections called "patches". Each of these patches becomes a vector that encapsulates the information relevant to that specific region of the spectrum. To preserve the original structure of the spectrum, positional coding is added to each patch. Subsequently, Vision Transformer uses transformer layers with a self-attenuating

mechanism, which captures long-range relationships between different regions of the spectrum, providing a global and detailed view of proteomic patterns. This approach allows modeling complex interactions within proteomic data, fundamental for predicting bacterial resistance.

3.1 Model used

The proposed convolutional vision transformer model (**Figure 2**) for bacterial resistance classification is structured in several blocks (X, Y, Z) with variable dimensions (64, 128, 256) and culminates in a fully connected layer. The architecture combines vision transformers and convolutional layers, which allows the capture of local and global features from the input images. At the end of the model, images are classified into two categories: sensitive or resistant.

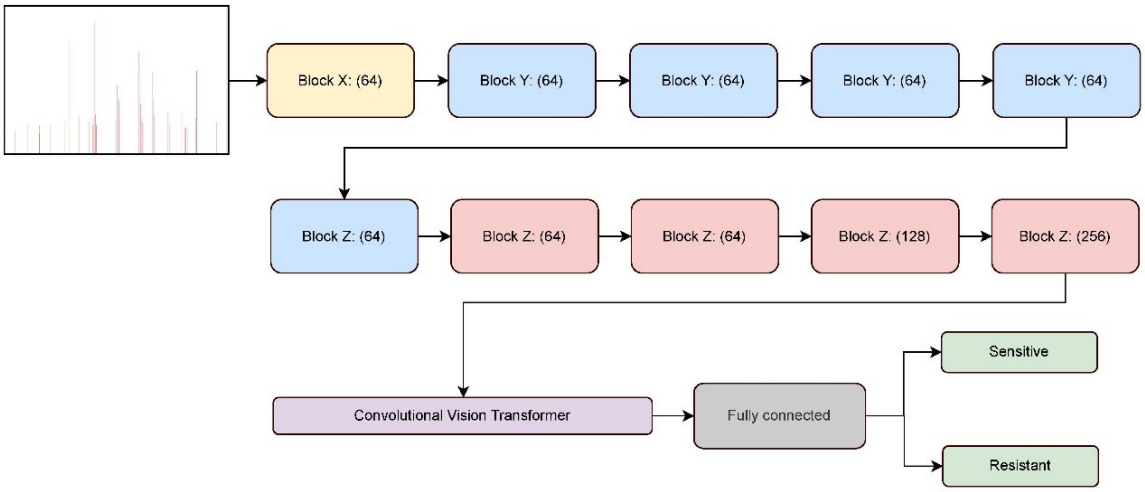

Figure 2: CVT architecture for bacterial resistance classification, consisting of convolutional blocks and transformer layers, culminating in a binary classification: sensitive or resistant.

- Block X **Figure 3** of the architecture consists of two 2D convolutional layers (Conv2D) with a kernel size of 3x3, a 1x1 step and "Same" padding, which ensures that the output dimensions are the same as the input dimensions. After each convolution, batch normalization is applied to stabilize and speed up the training process. The activation function used is Leaky ReLU, which allows a small gradient when the input is negative, which improves the gradient flow during training. This

block helps to extract local features from the input image, while optimizing model performance and stability.

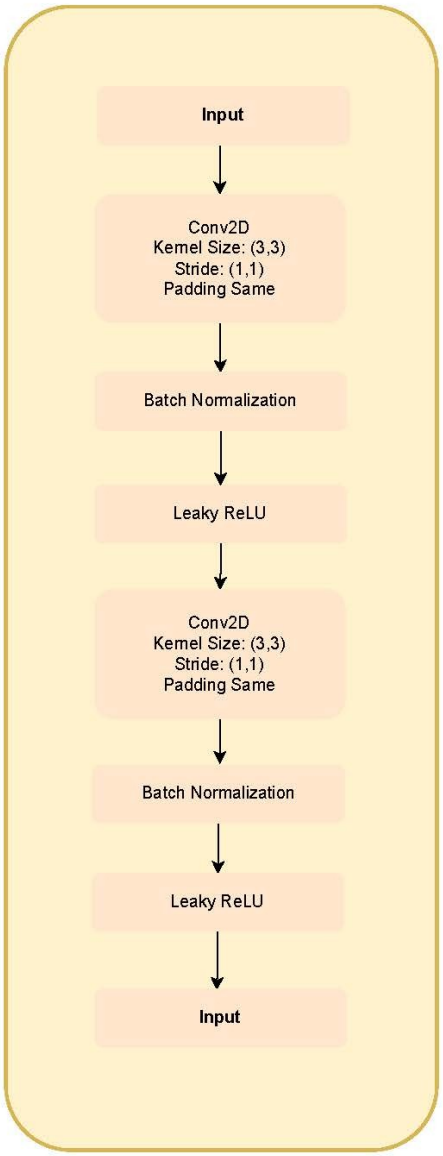

Figure 3: Block X: Two Conv2D layers with 3x3 kernel, 1x1 stride, 'Same' padding, followed by batch normalization and Leaky ReLU activation for local feature extraction and stable training

- The Y block **Figure 4** consists of two Conv2D layers with a kernel size of 3x3, a 1x1 step and "Same" padding, which ensures that the output dimensions are the same as the input dimensions. After each convolution, batch normalization is applied to stabilize the training process. After the second convolution, a compression and excitation (SE) block is included, which dynamically adjusts the importance of feature channels, improving the model's ability to focus on relevant features. Finally, the output of the AND block contains the processed features for the next stages of the model.

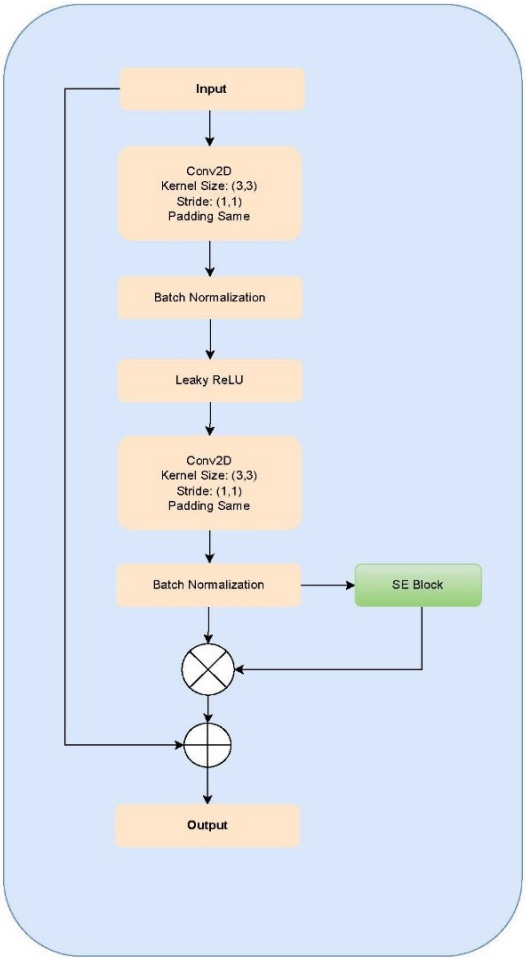

Figure 4: Block Y: Two Conv2D layers with 3x3 kernel, 1x1 stride, 'Same' padding, followed by batch normalization and a SE (Squeeze-and-Excitation) block for dynamic feature channel adjustment

- Block Z **Figure 5** consists of three Conv2D layers, each with a kernel size of 3x3, a 1x1 step and "Same" padding, which ensures that the output dimensions are the same as the input dimensions. After each convolution, batch normalization is applied to stabilize the training process. After the first convolution, Leaky ReLU activation is used to improve the gradient flow when the inputs are negative. Between the second and third convolution, an SE block is included that dynamically adjusts the importance of the feature channels, improving the model's ability to focus on the most relevant features. Finally, the block generates the processed features for the next stages of the model.

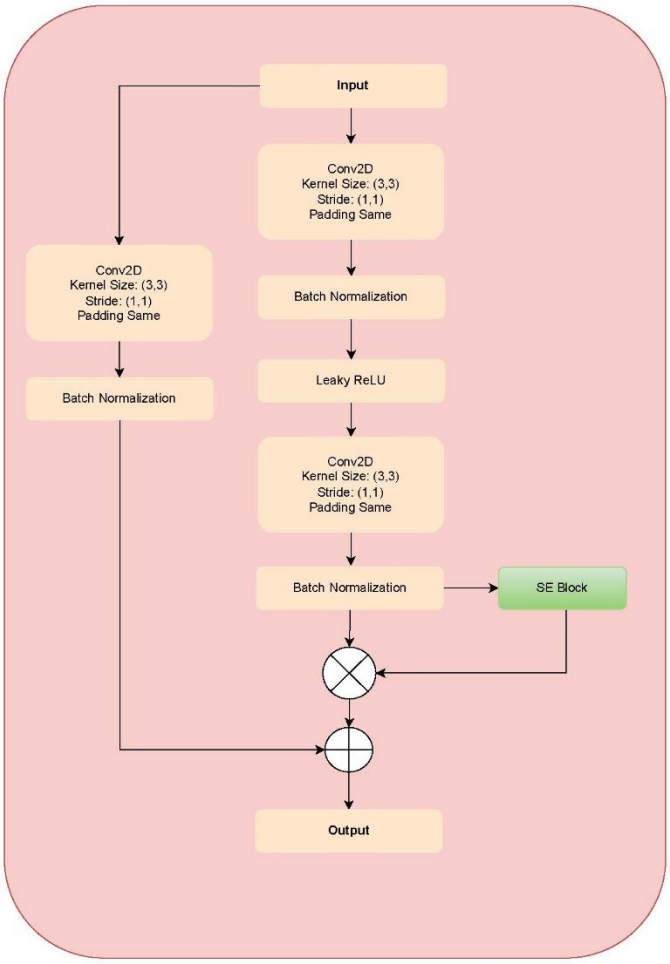

Figure 5: Block Z: Three Conv2D layers with 3x3 kernel, 1x1 stride, 'Same' padding, followed by batch normalization, Leaky ReLU activation, and a SE block for dynamic feature channel adjustment

- This architecture integrates a compression and excitation layer (SE-Block) **Figure 6** composed of a two-dimensional global average clustering layer, followed by a dense layer with Leaky ReLU activation and another dense layer with sigmoid activation. The SE-Block functions as a channel attention mechanism, allowing the model to autonomously evaluate the importance of each channel and adjust its weights during training. Convolutional models typically extract features by combining spatial information with information from each channel within the local receptive field. The main objective of SE-Block is to enhance the most relevant channels and decrease the influence of the less important ones. The compression operation consists of encoding a spatial feature into a global function, which is achieved by clustering global averages. This reduces the dimensionality of the feature channels, which increases the efficiency of subsequent computations. The excitation operation is applied to the global feature. This starts with a dense layer with Leaky ReLU activation, followed by a dense layer with sigmoid activation, adjusting the block dimensions. This recalibration of each channel's features helps the model focus on the features most critical to its task.

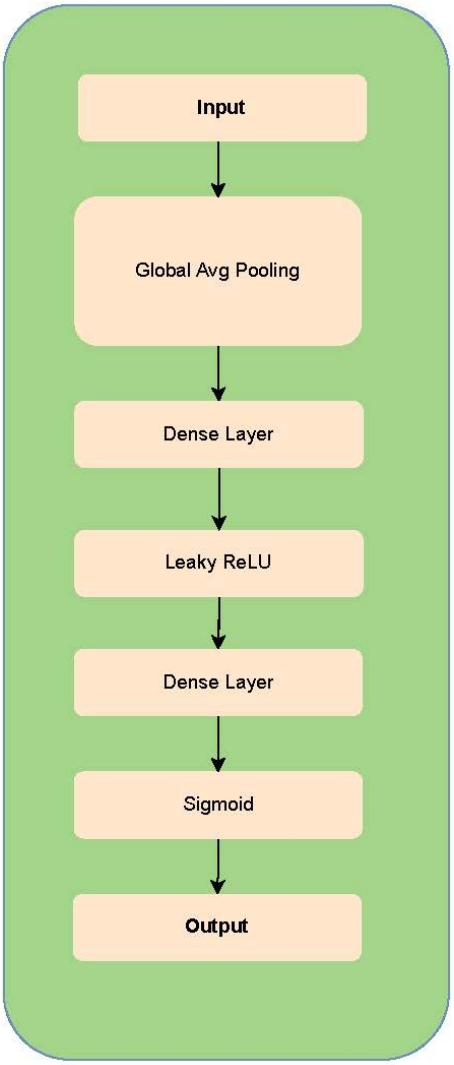

Figure 6: SE-Block: Composed of Global Average Pooling, two Dense layers with Leaky ReLU and Sigmoid activations, used to recalibrate channel-wise features by assigning dynamic weights to each channel.

- Convolutional Vision Transformers (CVT) have demonstrated excellent results in machine vision tasks in recent years, thanks to the combination of convolutional

networks and Vision Transformers. This fusion allows capturing and modeling both global and local features.

The CVT architecture used in this research is illustrated in **Figure 7**, where certain operations consolidate the feature maps, preserving the enhanced features from the previous blocks. The sequence starts with a 3x3 dimensional convolutional layer, followed by a Batch Normalization layer and a ReLU activation function with Leakage. Next, another 3x3 dimensional convolutional layer is added, along with another Batch Normalization layer. The processed information passes through a Se Block on the right-hand side before entering the transformer architecture. It should be noted that, unlike the previous Se Block, which used ReLU with Leakage, this block uses ReLU in its bottleneck.

Within the transformer architecture, each block starts with the convolutional embedding of 2D tokens, which allows the feature maps to undergo spatial subsampling, thus enriching the spatial representation of the tokens. The input tokens are reconfigured in 2D before being processed by the convolutional transformer block, where convolutional projections are applied before autoattenuation to obtain the query, key and value vectors ($Q=K=V$), instead of using positional linear projections as in Vision Transformers (ViT). The vectors are flattened to facilitate the calculation of multiheaded attention. The convolutional projection includes a multihead attention block and an MLP, with a normalization layer added before each component (Bravo-Ortiz et al., 2024; Holguin-Garcia et al., 2024).

Positional embedding is considered, since convolutional operations in token embedding and projection effectively capture spatial features from both global and local perspectives. The proposed network employs learnable one-dimensional positional embeddings, which are added to the token embeddings in the Convolutional Transformer. Finally, the convolutional projection is adjusted to modify both the feature dimension of each token and the number of tokens at each stage.

The proposed transformer block with convolutional projection extends the original transformer by incorporating local spatial context modeling and improving efficiency by allowing the undersampling of K and V matrices using depth separable convolutional layers to implement multihead self-attenuation. The projected tokens are finally compressed to a one-dimensional format for further processing.

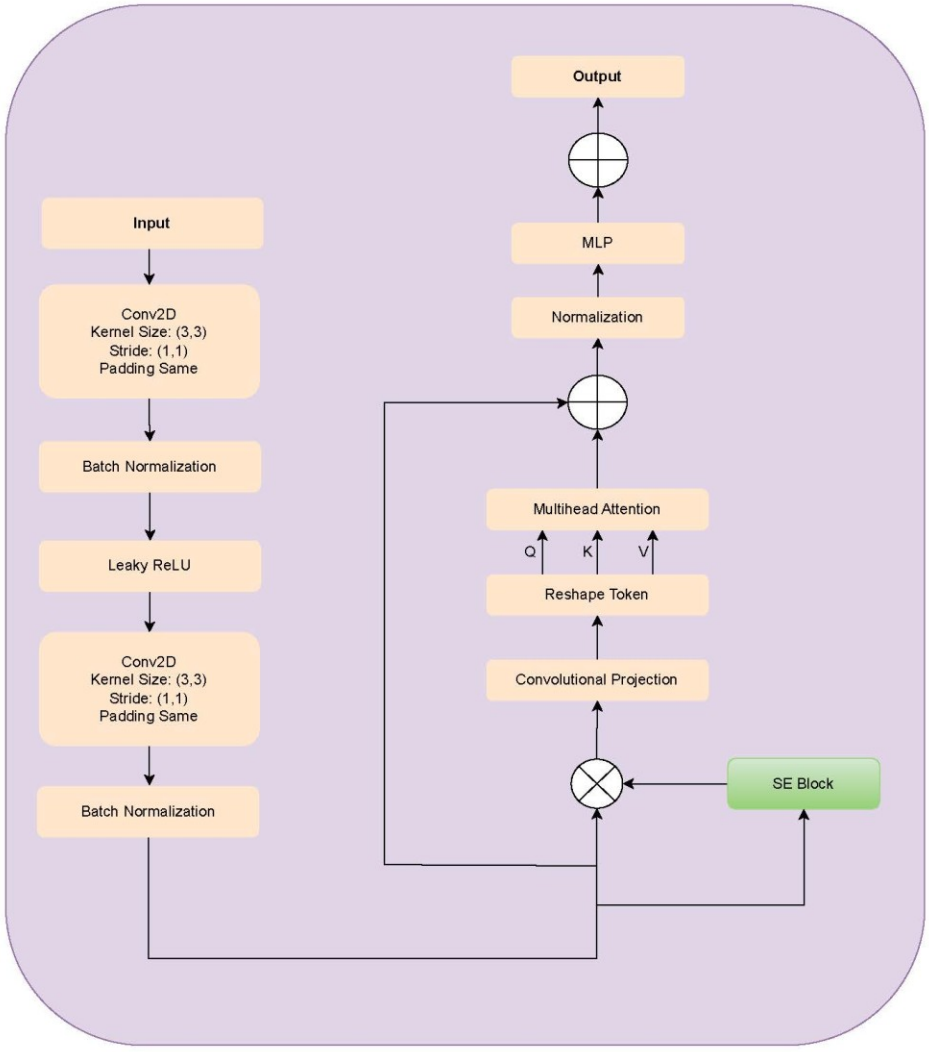

Figure 7: Convolutional Vision Transformer Architecture for Bacterial Resistance Classification with Squeeze-and-Excitation Block.

3.2 Parameters and Hyperparameters

3.2.1 Batch Normalization

Batch normalization is a technique used in deep learning to normalize the inputs of each layer to improve training speed, stability, and model performance. It helps to reduce the internal drift of covariates by keeping the distribution of activations consistent between layers (Luo et al., 2018.).

Batch normalization works as follows:

$$\boldsymbol{x}^{(\boldsymbol{i})} = \frac{\boldsymbol{x}^{(\boldsymbol{i})} - \boldsymbol{\mu B}}{\sqrt{\boldsymbol{\sigma}_{\boldsymbol{B}}^{\boldsymbol{2}} + \boldsymbol{\epsilon}}}$$

Where:

- $\boldsymbol{x}^{(() (\boldsymbol{i}) ())}$is the i-th entry in the batch
- $\boldsymbol{\mu}_{\boldsymbol{B}}$is the batch mean,
- $\boldsymbol{\sigma}_{\boldsymbol{B}}^{\boldsymbol{2}}$ is the variance of the lot,
- $\boldsymbol{\epsilon}$ is a small constant added to avoid division by zero.

After normalization, the normalized input $\mathbf{x}\text{^}(((() (\boldsymbol{i}) ()))$ is scaled and shifted by the learnable parameters $\boldsymbol{\gamma}$ and β**:**

$$y^{(i)} = yx^{\wedge(i)} + \beta$$

Where:

- $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are learned parameters used to scale and shift the normalized values.

In summary, batch normalization applies the following transformation:

$$y^{(i)} = \gamma \frac{x^{(i)-}\mu B}{\sqrt{\sigma^2 B + \epsilon}} + \beta$$

This helps stabilize and accelerate the training of deep neural networks by maintaining a controlled distribution of activations.

3.2.2 Global Average Pooling

The mathematical operation of the Global Average Pool (Hsiao et al., 2019) for each channel c is defined as:

$$y_c = \frac{1}{H \times W} \sum_{i=l}^{H} \sum_{j=l}^{w} x_{ijc}$$

Where:

- y_c is the average value of the channel c at the GAP output,
- $x_{(i)\,(jc)}$ is the value at position (i, j) in channel c of the input feature map,
- ***H** and **W*** are the height and width of the feature map, respectively.

This operation outputs a vector where each value represents the global significance of a feature channel across the entire spatial dimension.

3.2.3. Softmax

Softmax is a mathematical function frequently used in machine learning, especially in the output layer of classification models. It converts a vector of raw scores (logits) into probabilities, where each score represents the probability of belonging to a specific class. The Softmax function is particularly useful when dealing with multiclass classification problems, as it guarantees that the output values are positive and sum to 1 (Gold, 1996).
The Softmax function for a vector $\mathbf{z} = [z_1, z_2, \ldots, z_n]$ is defined as:

$$\boldsymbol{\sigma}(\boldsymbol{z_i}) = \frac{e^{\boldsymbol{zi}}}{\sum_{\boldsymbol{j=l}}^{\boldsymbol{n}} e^{\boldsymbol{zj}}}$$

Where:

- $\boldsymbol{\sigma(z_i)}$ is the probability associated with the i-th class,
- z_i represents the raw (logit) score of the i-th class,
- ***n*** is the total number of classes.

The Softmax function maps each zi score to a probability, guaranteeing that:

$$\sum_{i=l}^{n} \sigma(zi) = 1$$

This makes Softmax ideal for multi-class classification, as each output value represents the model's confidence in the assignment of a particular class.

3.2.4 Leaky ReLU

The Linear Leaky Rectified Unit (LEU) is a variant of the LEU activation function, which addresses the problem of "dying" ReLUs. In a standard ReLU function, any input less than zero is set to zero, which can cause gradients to vanish during training. LEU with Leakage mitigates this problem by allowing a small, non-zero gradient for negative inputs (J. Xu et al., 2020).

The mathematical definition of LEU with Leakage for an input ***x*** is:

$$Leaky\ ReLU(x) = \begin{cases} x, & if\ x \geq 0 \\ ax, & if\ x < 0 \end{cases}$$

Where:

- ***x*** is the input,
- **α** is a small constant (usually α = 0.01) that defines the slope for negative values.

This small slope for negative values helps to maintain a gradient, which allows the network to learn even when the input values are negative.

3.3 Transformer hyperparameters

Hyperparameter	Description	Value	Range	Step
LAYER NORM EPS	This hyperparameter is used in layer normalization. EPS (epsilon) is a small value added to prevent di- vision by zero during normalization.	1e-6	1e-5 - 1e-1	1e-n, n=1,2,...,6
NUM HEADS	This hyperparameter specifies the number of attention heads in the multi-head attention mechanism. Using multiple heads enables the model to attend to various parts of the input at the same time.	2	2 - 16	4
NUM LAYERS	This parameter specifies the number of layers in the network, particularly within the transformer block or attention block.	2	2 - 16	4
MLP UNITS	Defines the number of units in the multi-layer perceptron (MLP) network following the attention layer.	16	2 - 16	2
LEARNING RATE	This hyperparameter controls the learning rate of the optimizer, determining the size of adjustments made to the model weights during training.	1e-4	1e-6 - 1e-0,	1e-n, n=1,2,...,6
WEIGHT DECAY	This hyperparameter is used in the optimizer to help prevent overfitting. It applies a penalty to large weights during training, encouraging the model to keep weights smaller.	1e-3	1e-0 - 1e-6	1e-n, n=1,2,...,6

Table 1: Hyperparameters used in the architecture, along with a description of each hyperparameter, the range, and the step at which they were iterated for optimization

3.4 Transfer Learning

3.4.1 VGG16

The VGG-16 model is a convolutional neural network architecture created by the Visual Geometry Group (VGG) at the University of Oxford. This architecture starts with an input layer of dimensions 224x224x3. This is followed by a convolutional block containing two consecutive convolutional layers, each with 64 filters of size 33. The same padding is used to maintain spatial resolution. Next is a Maximum Clustering layer with a group size of 22 and a step size of 2.

The next block also has two convolutional layers, this time with 128 filters and a filter size of 33, followed by another Maximum Clustering layer with a group size of 22 and a step size of 2. The architecture includes two consecutive convolutional layers with 256 filters each, using a filter size of 33. Next, two blocks are placed, each with three convolutional layers with 512 filters of size 33. After each of these blocks, a Maximum Clustering layer with a group size of 22 and a step size of 2 is placed. The network concludes with three fully connected layers: the first layer contains 25088 elements and generates 4096, the second layer also generates 4096 and the final layer generates 1000, representing the 1000 classes of the ILSVRC challenge. In the final layer a Softmax activation function is used for classification (Simonyan & Zisserman, 2015).

3.4.2 InceptionResNetV2

The InceptionResNetV2 model is a deep convolutional neural network that combines the powerful features of the Inception and ResNet architectures, providing high accuracy in image classification while maintaining computational efficiency. With Inception modules, it effectively captures detailed and broader patterns, and ResNet's residual connections help improve gradient flow, enabling stable training in very deep layers. This hybrid design allows InceptionResNetV2 to recognize complex patterns with a reduced number of parameters compared to other large networks, making it ideal for tasks that require high accuracy and resource efficiency (Dash et al., 2023).

3.4.3 DenseNet201

The DenseNet201 architecture is based on a series of dense blocks, each with multiple convolutional layers. A distinguishing feature of DenseNet201 is the use of dense connections, where each dense block incorporates as inputs the outputs not only of the immediately preceding layer, but also of all previous layers within the block. This structure allows the output of each layer to be concatenated and used by subsequent layers, ensuring a complete flow of information throughout the network. To manage dimensionality and the number of feature maps, transition layers are placed between dense blocks. These layers include batch normalization, a 1x1 convolution to reduce feature maps, and a clustering layer to reduce spatial dimensions. With a depth of 201 layers, DenseNet201 has significantly more parameters than DenseNet121, which nearly doubles its complexity and computational demand, resulting in a more resource-consuming model for both training and inference (G. Huang et al., 2017)

3.4.4 EfficientNetB7

The EfficientNetB7 architecture consists of 813 layers organized into 5 main modules, each contributing to specific functions within the network. Module 1 includes Conv2D DepthWise, Batch Normalization and an activation function, laying the foundation for the sub-blocks. Module 2 consists of Conv2D DepthWise, Batch Normalization, activation, zero padding, another Conv2D DepthWise and Batch Normalization, marking the start of the first sub-block in all but the first seven main blocks. Module 3 contains Global Average Clustering, a Recovery function and two Conv2D layers, which function as a jump connection between all sub-blocks. Module 4 integrates Multiplication, Conv2D and Batch Normalization, and is used to merge jump connections in the initial sub-blocks. Module 5, which combines Multiplication, Conv2D, Batch Normalization and Dropout, links each sub-block to the previous one via a jump connection, effectively merging them. These modules are grouped into specific sub-blocks that are used within the main blocks of the architecture. Sub-block 1 is used only as the initial sub-block of the first block, sub-block 2 acts as the first sub-block in all other blocks, and sub-block 3 is applied to all sub-blocks except the first one in each block (Tan & Le, 2020.).

3.4.5. EfficientNetV2L

EfficientNetV2L is a highly optimized convolutional neural network architecture designed for fast and accurate image classification. Based on the EfficientNet family, it combines enhanced model scaling with advanced training techniques to achieve superior performance with lower computational costs. EfficientNetV2L uses a composite scaling approach, balancing depth, width and resolution, and introduces fused MBConv layers to improve efficiency. This architecture is known for its reduced training time and high accuracy, making it ideal for real-time applications and tasks that require efficient implementation on large datasets (Tan & Le, 2021).

3.4.6 ConvNeXtXLarge

ConvNeXtXtXLarge is a convolutional neural network model designed to take advantage of recent advances in computer vision, pushing the capabilities of traditional ConvNets to a new level. Launching in 2022, ConvNeXtXtXLarge incorporates deep, scalable convolutional blocks, extensive use of normalization and activation layers, and modern scaling techniques. These elements emphasize that ConvNets can not only match, but even exceed, the performance of newer architectures such as Transformers in vision tasks. Through optimized design and tuning, ConvNeXtXLarge achieves computational efficiency while delivering competitive results in standard benchmarks such as ImageNet. Its ability to handle high-resolution images demonstrates its versatility in various vision applications (Liu et al., 2022).

3.5 Fully Connected

The fully connected layer responsible for classifying convolutional models in transfer learning starts with dense 128, 64 and 32 neurons, with a ReLU activation layer and batch normalization.

3.6 Metrics

Tabares-Soto et al. (2021) highlight the importance of using metrics in model evaluation, emphasizing the importance of distinguishing between false positives (FP), false negatives (FN), true positives (VP) and true negatives (VN).
The following are the definitions and applications of the most important metrics.

3.6.1 Accuracy

Accuracy is a measure ranging from 0 to 1 and represents the percentage of correct predictions made by the model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

3.6.2 Precision

Accuracy indicates the proportion of true positives over total positive predictions, which helps to identify the class distribution accurately.

$$Precision = \frac{TP}{TP + FP}$$

3.6.3 Sensitivity (Recall)

Also known as sensitivity, recall shows the ability of the classifier to correctly identify positive instances in the data set.

$$Recall = \frac{TP}{TP + FN}$$

3.6.4 F1 Score

The F1 score is a metric that balances accuracy and recovery, and is useful for addressing class imbalance.

$$F1 = 2 * \frac{Presicion * Recall}{Presicion + Recall}$$

3.6.5 Support

The bracket indicates the number of instances in each test class, which provides an idea of the size of the classes and their representatives in the dataset.

3.6.6 Confusion Matrix

The confusion matrix is a tool that shows the relationship between the actual classes and those predicted by the model. The rows represent the predicted classes, while the columns represent the actual classes.

3.6.7 Cross-Validation

Cross-validation, a technique used to evaluate model performance, is a practical and widely used method. It divides the data into several subsets called "folds" (k) of similar sizes, allowing us to obtain multiple training and test sets. This helps us to better understand model performance and generalization.

$$Cross - Validation = \frac{1}{k} \sum_{i=l}^{k} Performance$$

These metrics provide a holistic understanding of the model's performance, allowing us to evaluate its accuracy, generalization and handling of class imbalances.

3.7 Hardware and resources

Google Colab was used for all experiments. Two different GPUs were used: an NVIDIA T4 GPU with 15 GB of memory, CUDA version 10.1 and 12 GB of RAM for general experiments, and an NVIDIA A100 GPU with 40 GB of memory for EfficientNetB7, DenseNet201 and ConvNeXtXLarge, which provided superior performance with enhanced capabilities. TensorFlow 2.15.0 and Python 3.10.12 were used for all tests

Results

1.Results and performance analysis of Machine Learning models

1.1 Evaluation through cross-validation

Fourteen supervised learning models were evaluated using 10-fold stratified cross-validation to classify MALDI-TOF spectra of bacteria as carbapenem-resistant or carbapenem-resistant (KPC). The metrics analyzed were accuracy, precision, recall and F1-score (**Table 2**).

	Performance metrics			
Models	**Accuracy**	**Precision**	**Recall**	**F1-score**
Logistic regression	61%	40%	13%	19%
Random forest	61%	40%	13%	19%
Suppor Vector Machine	61%	40%	13%	19%
K-Nearest Neighbors	45%	45%	100%	62%
Naive Bayes classifiers	61%	40%	13%	19%
Gradient boosting	61%	40%	13%	19%
LightGBM	55%	0%	0%	0%
Decision Tree Classifier	61%	40%	13%	19%
Naive Bayes Multinomial	58%	40%	7%	12%
XGBoost Classifier	56%	20%	2%	4%
Quadratic Discriminant Analysis	55%	0%	0%	0%
AdaBoost Classifier	61%	40%	13%	19%
Extra Tree Classifier	61%	40%	13%	19%

Table 2. Evaluation of model performance by cross-validation

Although most of the models (Random Forest, SVM, Logistic Regression, AdaBoost, Gradient Boosting, Extra Trees, among others) achieved an overall accuracy close to 61%, their recall and F1-score values were considerably low, ranging between 0% and 19%. This situation reflects that the models, although achieving an acceptable overall classification in terms of majority class, are not being able to correctly identify the positive class (resistant bacteria), which represents a significant risk in the clinical context.

The K-Nearest Neighbors (KNN) model was the only one that managed to break this trend, achieving a recall of 100% and an F1-score of 62%, although with a reduction in accuracy

(45%). This indicates that KNN, unlike the rest, does manage to identify all positive cases, although it incurs more false positives, a compromise that could be acceptable in clinical scenarios where the priority is not to miss resistant infections. In contrast, models such as LightGBM, XGBoost and QDA collapsed completely, with recall and F1-score metrics equal to zero. These results evidence a structural problem related to class imbalance and the low discriminative ability of the current attributes.

1.2 Evaluation by confusion matrix

After cross-validation evaluation, the final performance of each model was measured using the confusion matrix on an independent test set. This matrix allows direct observation of the distribution of true positives, false negatives, true negatives and false positives, as well as the calculation of more intuitive metrics for clinical decision making (**Table 3**).

		Performance metrics		
Models	**Accuracy**	**Precision**	**Recall**	**F1-score**
Logistic regression	62%	100%	13%	22%
Random forest	62%	100%	13%	22%
Suppor Vector Machine	62%	100%	13%	22%
K-Nearest Neighbors	62%	100%	13%	22%
Naive Bayes classifiers	62%	100%	13%	22%
Gradient boosting	62%	100%	13%	22%
LightGBM	56%	0%	0%	0%
Decision Tree Classifier	62%	100%	13%	22%
Naive Bayes Multinomial	59%	100%	6%	11%
XGBoost Classifier	59%	100%	6%	11%
Quadratic Discriminant Analysis	56%	0%	0%	0%
AdaBoost Classifier	62%	100%	13%	22%
Extra Tree Classifier	62%	100%	13%	22%

Table 3. Evaluation of the models by confusion matrix

The results confirmed the trend observed in the cross-validation. Most of the models (Logistic Regression, SVM, Random Forest, AdaBoost, Gradient Boosting, Extra Trees, among others) showed an *accuracy* of 100%, but with a very low *recall* of 13%, again indicating a strong tendency not to predict the positive class. This resulted in an *F1-score* of 22%, consistent with previous validation. Models such as LightGBM and QDA collapsed

completely with metrics close to 0%, while XGBoost and Naive Bayes Multinomial achieved a recall of just 6% and F1-scores of less than 11%.

It is worth noting that, unlike what was observed in the cross-validation, the **KNN** model did not stand out in this evaluation with confusion matrix (**Figure 8**), presenting identical metrics to the rest of the models. This can be attributed to the sensitivity of KNN to the distribution of the specific test set, which evidences the importance of considering multiple partitions (as in CV) to avoid over-interpreting the results in a single run. Overall, this section confirms that while some models are useful for excluding negative cases, their ability to identify resistant bacteria remains insufficient without applying additional adjustments such as class balancing or alternative approaches such as image-based.

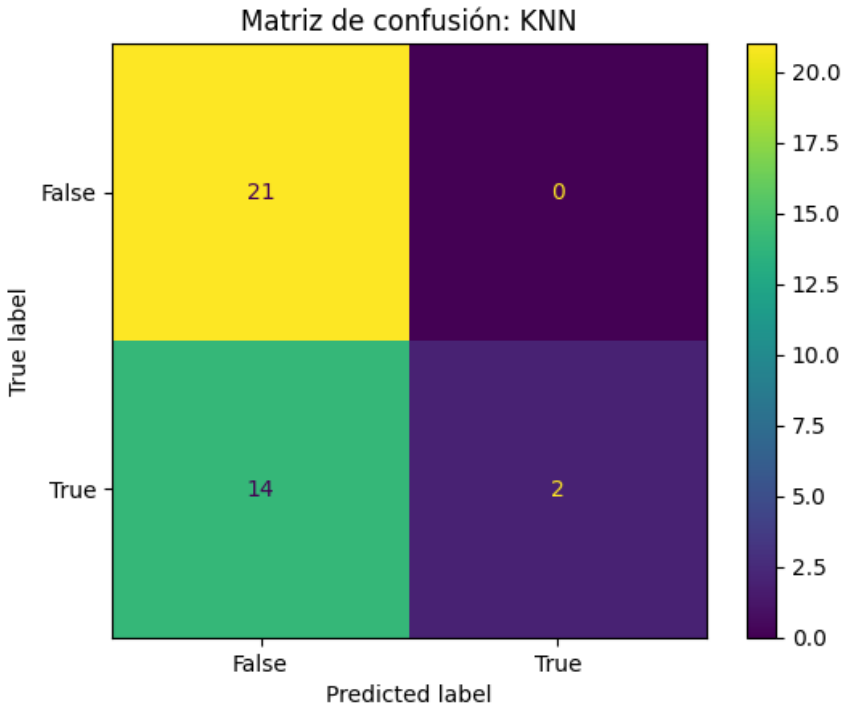

Figure 8. Evaluation by confusion matrix of the KNN model.

To address this potential limitation and improve the ability to detect bacterial resistance, an approach based on spectral imaging and Deep Learning was explored using the Convolutional Vision Transformer (CVT) architecture. This strategy leverages the potential

of visual transformational models, which have demonstrated great ability to extract complex and subtle spatial patterns in graphical representations of high-dimensional data. The conversion of spectra to images allows capturing relationships between regions of the spectrum that might be lost in traditional tabular representations, which, in conjunction with the contextual attention power of CVT, offers a promising avenue for improving diagnostic performance, especially in the identification of resistant bacteria with atypical or marginal patterns.

2. Performance of the Convolutional Vision Transformer Model

Table 4 presents the results obtained in the investigation, where transfer learning models were used to evaluate the performance of the proposed architecture. The VGG16 model shows an average accuracy of 70.91 % with a high standard deviation of ±12.71 %, indicating considerable inter-fold instability. Its precision (53.68 %), recall (70.91 %) and F1 score (60.66 %) metrics are low and exhibit high variability, suggesting that VGG16 is not reliable for this task. InceptionResNetV2 improves the average accuracy, reaching 70.12 % with a standard deviation of ±6.14 %, indicating higher inter-fold consistency. Its average values for precision (70.16%), recall (70.12%) and F1 score (70.14%) are slightly higher, demonstrating a more balanced performance than VGG16.

DenseNet201 shows better performance, with an average accuracy of 72.41 % and a lower variability of ±6.27 %. Its accuracy (72.41 %), recall (72.41 %) and F1 score (72.41 %) metrics are consistent, making it a more robust model for this task compared to previous models. The EfficientNetB7 model performs similarly to DenseNet201, with an average accuracy of 72.41 % and a standard deviation of ±6.27 %, although it has slight variations in precision and recall, suggesting slightly less consistent performance.

EfficientNetV2L stands out for its high average accuracy of 75.85% and low standard deviation of ±3.91%, indicating reliable and stable performance. Its 75.86% accuracy, 75.85% recovery and 75.85% F1 score metrics are well aligned, indicating a good classification capability for this task. ConvNeXtXtXLLarge also demonstrates remarkable

performance with an average accuracy of 75.85% and a similar standard deviation to EfficientNetV2L. Its 75.86% accuracy, 75.85% recovery and 75.85% F1 score metrics are consistent, making it competitive with EfficientNetV2L.

Finally, the CVT model achieves the best performance, with an average accuracy of 80.19% and a standard deviation of ±6.17%, reflecting a balance between high accuracy and consistency. Its accuracy (79.41 %), recovery (80.19 %) and F1 score (79.79 %) metrics are also the highest among the evaluated models, indicating that CVT is the most suitable model for this task, outperforming the others in terms of effectiveness and classification stability.

Model	Metric	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean
VGG16	Accuracy	58,62%	58,62%	89,65%	65,51%	82,14%	70.91%± 12.71%.
	Precision	34,36%	58,62%	89,77%	69,19%	81,91%	66.77%± 19.39 %.
	Recall	58,62%	58,62%	89,65%	65,51%	82,14%	70.91%± 12.71%.
	F1-score	43,32%	58,62%	89,60%	64,24%	81,91%	67.54%± 16.55%.
InceptionResNetV2	Accuracy	65,51%	68,96%	82,75%	65,51%	67,85%	70.12%± 6.40%.
	Precision	66,63%	70,45%	83,90%	65,98%	77,34%	72.86%± 6.84%.
	Recall	65,51%	68,96%	82,75%	65,51%	67,85%	70.12%± 6.45%.
	F1-score	65,76%	68,96%	82,37%	65,43%	68,14%	70.13%± 6.26%.
DenseNet201	Accuracy	72,41%	65,51%	68,96%	65,51%	75,00%	69.48%± 3.76%.
	Precision	72,15%	65,74%	68,83%	66,46%	75,66%	69.77%± 3.70%.
	Recall	72,41%	65,51%	68,96%	65,51%	75,00%	69.48%± 3.76%.
	F1-score	71,92%	64,29%	68,81%	64,67%	75,23%	68.98%± 4.20%.
EfficientNetB7	Accuracy	72,41%	58,62%	72,41%	62,06%	67,85%	66.67%± 5.53%.
	Precision	78,57%	76,35%	73,48%	62,88%	70,65%	72.39%± 5.45%.
	Recall	72,41%	58,62%	72,41%	62,06%	67,85%	66.67%± 5.53%.
	F1-score	72,21%	46,52%	71,43%	61,79%	68,49%	64.09%± 9.51%.
EfficientNetV2L	Accuracy	65,51%	75,86%	68,96%	62,06%	75,00%	69.48%± 5.33%.
	Precision	68,36%	75,81%	69,12%	64,15%	77,69%	71.03%± 5.00%.
	Recall	65,51%	75,86%	68,96%	62,06%	75,00%	69.48%± 5.33%.
	F1-score	65,68%	75,74%	68,27%	61,14%	75,49%	69.26%± 5.66%.
ConvNeXtXLarge	Accuracy	65,51%	55,17%	65,51%	62,06%	75,00%	64.65%± 6.40%.
	Precision	66,63%	54,35%	65,51%	66,37%	75,66%	65.71%± 6.77%.
	Recall	65,51%	55,17%	65,51%	62,06%	75,00%	64.65%± 6.40%.
	F1-score	65,76%	54,17%	65,51%	60,08%	75,23%	64.15%± 6.97%.
CVT	Accuracy	72,41%	79,31%	86,20%	86,03%	89,28%	80.61%± 6.29%.
	Precision	75,46%	84,95%	86,72%	75,92%	90,81%	82.77%± 6.08%.
	Recall	72,41%	79,31%	86,20%	75,86%	89,28%	80.61%± 6.29%.
	F1-score	72,54%	77,84%	86,03%	75,80%	88,75%	80.19%± 6.17%.

Table 4: Evaluation of Accuracy, Precision, Recall, and F1-score for Various Deep Learning Models on Five-Fold Cross-Validation.

To complete the evaluation of the proposed model, its additional metrics by class and confusion matrix can be seen in **Figure 9.** In this case, the metrics are somewhat more

unstable, probably due to the slight imbalance in the data; however, the target class, "resistance", shows the highest results. In addition, Grad-CAM in **Figure 10** is applied to the processed image, which clearly shows how the model segments the most important parts for classification, while simply avoiding the rest, which is considered noise.

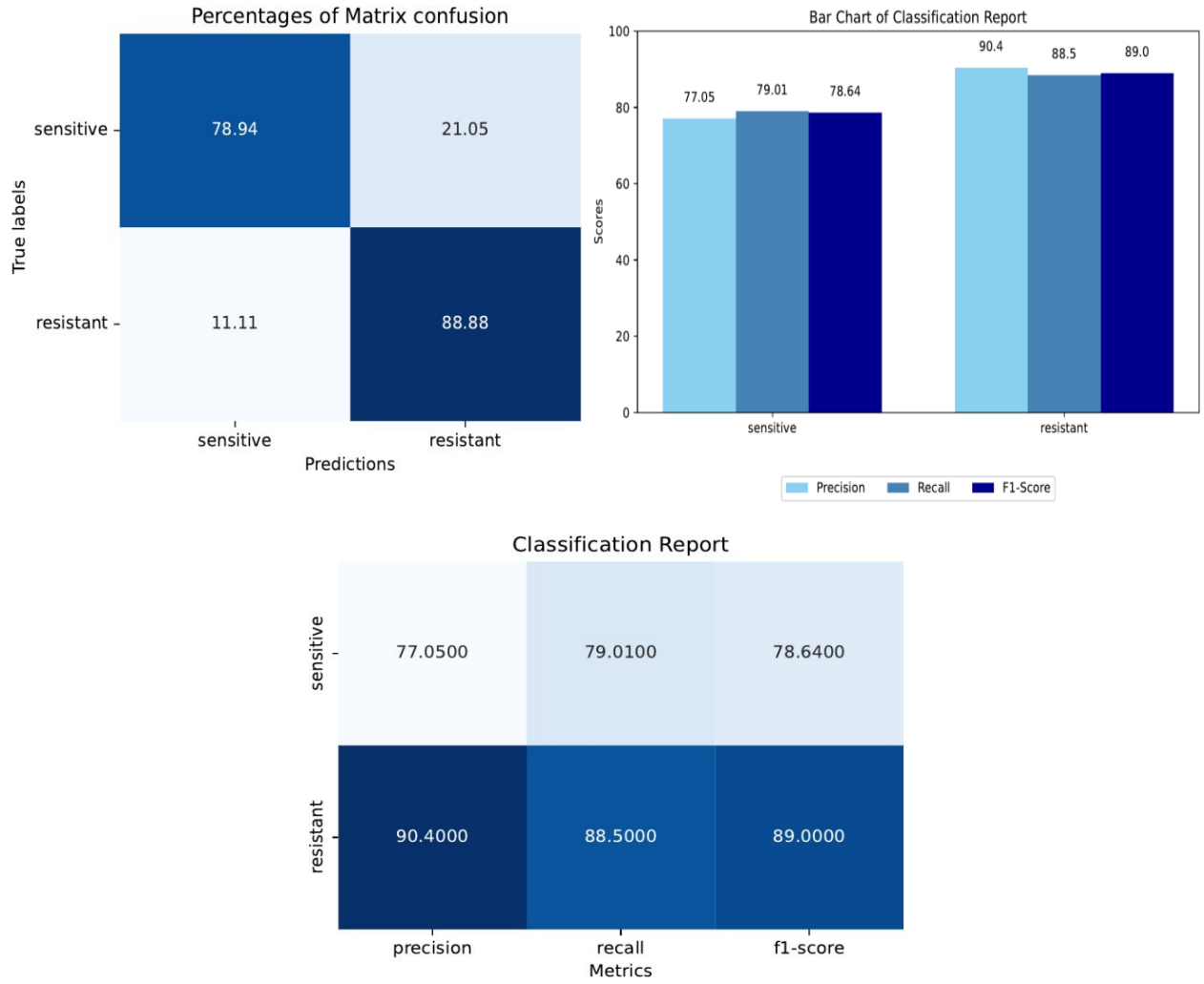

Figure 9: Confusion Matrix and Classification Report Showing Precision, Recall, and F1-score for Sensitive and Resistant Classes

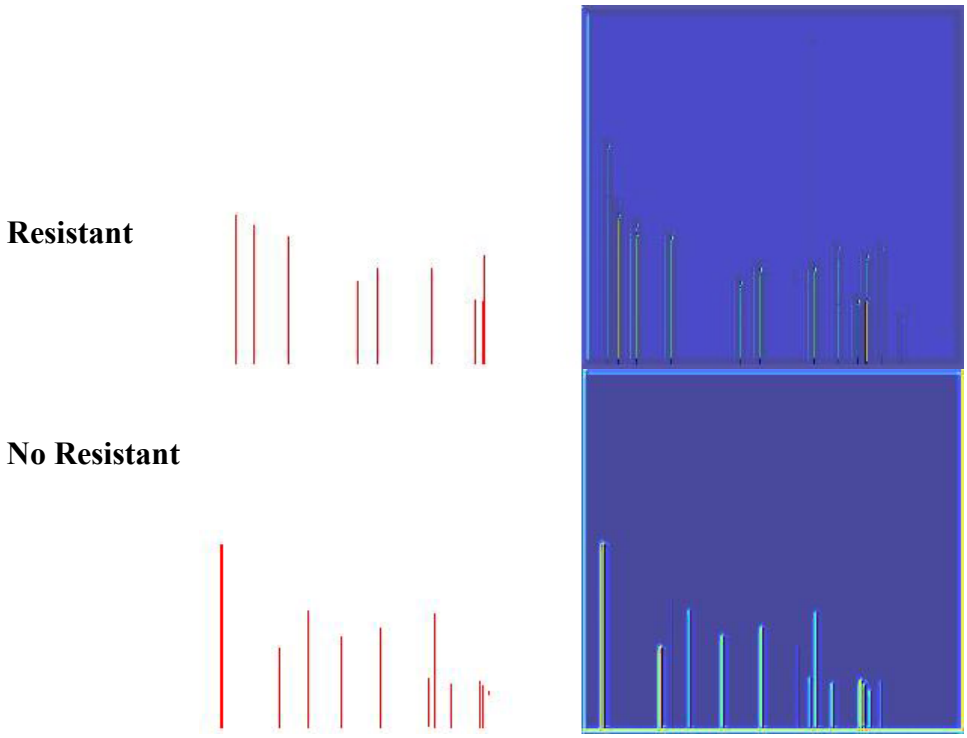

Figure 10: Grad-CAM visualization of the CVT model for bacterial resistance classification.

Discussion

Antimicrobial resistance is one of the greatest threats to public health, with a significant impact on morbidity, mortality and healthcare costs associated with bacterial infections. Rapid and accurate detection of resistant strains is crucial to guide appropriate therapeutic strategies and prevent their spread. However, conventional phenotypic detection methods, such as broth microdilution and agar diffusion, require 24-72 hours to obtain results, which limits their effectiveness in clinical settings, where rapid decision making is essential for effective infection management and mortality reduction.

In this context, MALDI-TOF mass spectrometry has revolutionized clinical microbiology by allowing the identification of microorganisms in a matter of minutes through the analysis of unique proteomic profiles. This represents a significant improvement over traditional method. Moreover, the integration of artificial intelligence (AI) with MALDI-TOF mass

spectrometry has further expanded its potential, allowing not only the identification of bacterial species, but also the prediction of antibiotic resistance patterns (Gato et al., 2021).

AI, particularly through machine learning and deep learning algorithms, has demonstrated an exceptional ability to process large volumes of proteomic data generated by MALDI-TOF mass spectrometry and detect complex patterns associated with bacterial resistance. For example, recent studies have shown that support vector machine (SVM), K-nearest neighbor (KNN) and random forest models, combined with MALDI-TOF mass spectrometry, have achieved accuracy rates of up to 97.83 % and F1 scores of 96.85 % in detecting carbapenem-resistant strains (Gato et al., 2023). These results exceed the efficiency of conventional methods by providing accurate predictions in significantly less time.

The present investigation aimed to evaluate the ability of different artificial intelligence models to predict bacterial resistance from spectra generated by MALDI-TOF mass spectrometry for carbapenem-resistant *K. pneumoniae*. We initially worked with classical machine learning models applied to numerical variables (m/z and intensity), but these showed limited performance. Although they achieved an overall accuracy close to 61 %, their recall and F1-score values were critically low (between 0 % and 19 %), reflecting a high false negative rate. This trend was confirmed in the confusion matrix, where models such as SVM, Random Forest and Logistic Regression presented recall of only 13 % and an F1-score of 22 %, evidencing their inability to identify the resistant class reliably. Although the KNN model achieved 100 % recall in cross-validation, this performance was not replicated in the final confusion matrix, suggesting a high sensitivity to data partitioning. These shortcomings indicated the need for an alternative methodological approach.

Faced with these limitations, it was proposed to treat the spectra as images and to use a hybrid architecture *of* Vision Transformer with convolutions (Convolutional Vision Transformer, CVT). This approach succeeded in capturing more complex and discriminative patterns, both local and global, which resulted in a considerable improvement in performance. The CVT model achieved an average accuracy of 80.19% with a standard deviation of ±6.17%, far outperforming traditional models. Its confusion matrix showed a balanced classification

capability: it correctly identified 88.88 % of resistant strains and 78.94 % of sensitive strains. These figures are reflected in the specific metrics of the classification report, with an accuracy of 90.4 %, recall of 88.5 % and F1-score of 89 % for the resistant class, positioning CVT as the most efficient and stable model of those evaluated. In addition, the use of interpretability techniques such as Grad-CAM allowed us to visualize that the most active regions during prediction coincided with the most relevant peaks of the spectrum, providing evidence of the model's ability to focus on informative spectral features. Taken together, these results not only confirm the initial hypothesis about the potential of AI in this task, but also highlight the superiority of the image-based approach and specialized architectures such as CVT for bacterial resistance prediction, with direct implications for the development of clinical diagnostic support tools.

The efficacy of this model is especially relevant in the fight against antimicrobial resistance, where rapid and accurate diagnosis is crucial to prevent the spread of resistant strains and improve patient outcomes. In addition, the use of Grad-CAM improved model interpretability by highlighting the most critical regions in the proteomic spectra that influenced classification decisions. This opens the door for future research investigating the bacterial proteome for phenotypic and genotypic relationships in their resistance mechanisms.

Comparing these results with recent studies, it is observed that the performance of the CVT model is in a competitive range against other machine and deep learning approaches investigating resistance prediction in *K. pneumoniae* against carbapenemics. For example,Huang et al., (2020) used models such as Random Forest and SVM on MALDI-TOF spectra, with 97% accuracy, albeit on a smaller and more homogeneous data set. Xu, (2024) reported accuracies of 84% (Random Forest), 81 % (SVM and logistic regression) and 85 % (XGBoost). Another study tested LASSO, Logistic Regression, SVM and Neural Networks (NN), obtaining test set accuracies of 87 %, 79 %, 62 % and 68 %, respectively (Zeng et al., 2023).

Another particularly interesting approach was that of Yu et al. (2023), who simultaneously explored resistance to carbapenemics and colistin in more than 2,000 clinical isolates of *K. pneumoniae*. To do so, they combined multiple data sources, including resistance genes, spectral profiling by MALDI-TOF, sensitivity tests, and a phenotypic test (NG-Test CARBA 5). Using the LightGBM model, they achieved an area under the curve (AUC) of 0.95 for carbapenem-resistant and 0.88 for carbapenem-resistant strains, and AUCs of 0.83 and 0.84 for colistin-resistant and colistin-sensitive strains, respectively. These results show the potential of combining different sources of information to achieve a robust and multifaceted diagnosis.

In the field of deep learning, two relevant contributions stand out; a study that analyzed 2,683 proteomic spectra of *K. pneumoniae* developed an artificial neural network (ANN), achieving consistent performance metrics: sensitivity, specificity, accuracy and an F1 score of 84 % (Zhang et al., 2023). Similarly, the MSDeepAMR model proposed by López-Cortés et al., (2024), based on deep neural networks on raw spectra, reported AUC values above 0.83 in predicting resistance in *E. coli*, *K. pneumoniae* and *S. aureus*. Although effective, our approach suggests that treating spectra as images brings additional advantages in capturing complex spatial relationships.

Regarding the use of genomic data, Condorelli et al., 2024) achieved accuracies of over 90% in the prediction of resistance to 10 antibiotics using genomic sequences of *K. pneumoniae.* However, these methods usually require greater investment of time and resources, which limits their applicability in clinical contexts that demand diagnostic immediacy.

Despite these advances, no previous studies have applied CVT architectures to proteomic data for bacterial resistance detection, highlighting the novelty of our approach. However, transformer-based models have been successfully implemented in related areas. For example, a study using the D1 transformer predicted bacterial resistance from hospital electronic medical records, incorporating multiple clinical factors such as previous antibiotic exposure, resistance history, hospital admission, age, and general clinical status (Tharmakulasingam et al., 2023). These findings demonstrate the potential of transformers in clinical applications

and suggest that their use in proteomic data analysis could generate new advances in resistance detection.

However, the limitations of the present study should be acknowledged. One of the main limitations observed in this study was the variability in the performance metrics of certain models, which can be attributed to several factors. First, the moderate imbalance between classes (more sensitive than resistant strains), as well as the overall sample size, could have negatively affected model performance. A balanced and robust data set is crucial to obtain reliable performance metrics and ensure that models can be effectively generalized to new data. Addressing this problem in future research could improve the stability and generalizability of AI models.

In addition, the nature of the data set used could have contributed to the variability in performance. Since the data were pre-existing, rather than from a controlled experimental system, some critical parameters that could have optimized data quality were not standardized. Variables such as choice of culture medium, incubation time, and precise calibration of the equipment could have influenced the generation of proteomic spectra, introducing noise and variability between samples.

In addition, the lack of preprocessing techniques to reduce noise, especially in the early and late regions of the spectra, could have introduced irrelevant signals, which hinders accurate pattern recognition. Implementation of advanced preprocessing techniques, such as noise reduction and filtering of uninformative signals, could improve model performance by improving signal clarity.

On the other hand, other limitations of this approach are the unclear association between spectral peaks and specific resistance mechanisms, since these models can identify discriminatory patterns without necessarily revealing the molecular basis of resistance. In addition, the high similarity between spectra generated by sensitive and resistant strains of the same species makes it difficult to differentiate between them, especially in the absence of clearly defined biomarkers.

To improve its applicability in clinical settings, the incorporation of external data sets to assess the impact of clonal variability is suggested, as well as the development of regional models with comprehensive validations, especially if this methodology is to be used as a single test for resistance detection in established bacteria (Kong et al., 2022).

Finally, MALDI-TOF is not a substitute for phenotypic antimicrobial susceptibility testing. However, the integration of MALDI-TOF with deep learning models, particularly with hybrid architectures such as CVT, represents a promising alternative. By treating spectra as images, these models can learn hierarchical representations and capture subtle patterns in peak distribution, enabling more robust classification between sensitive and resistant strains, even in the presence of noise or experimental variability.

Despite these limitations, the integration of MALDI-TOF mass spectrometry with Vision Transformers presents a promising approach for the detection of antimicrobial resistance. While this study focused on a single bacterial species and resistance to one type of antibiotic, the methodology could be extended to the detection of resistance in other pathogens, paving the way for further optimization of microbiological diagnosis and clinical decision making.

Conclusion

Bacterial resistance is one of the greatest challenges to global public health, with a significant impact on mortality, morbidity and costs associated with infection management. Addressing this problem requires innovative solutions that enable faster and more effective diagnosis. In this context, the integration of MALDI-TOF mass spectrometry with artificial intelligence algorithms has become a promising approach for early identification of bacterial resistance, offering significantly shorter diagnostic time compared to traditional methods.

Previous studies have demonstrated the value of these technologies, which have successfully classified sensitive and resistant strains with high accuracy, positioning them as essential diagnostic tools for clinicians. In particular, this study highlights the Convolutional Vision Transformer (CVT) as a key innovation, as no previous research has been found that applies

this architecture to bacterial resistance prediction based on MALDI-TOF proteomic spectra. The outstanding performance of CVT underscores its potential for early detection of carbapenem-resistant *Klebsiella pneumoniae* even before traditional antimicrobial susceptibility testing is completed.

The ability to significantly reduce the time to diagnosis has direct clinical implications, allowing timely therapeutic interventions and contributing to the reduction of mortality and morbidity associated with antimicrobial resistance.Therefore, the integration of CVT and MALDI-TOF MS not only improves patient care, but also represents a substantial advance in the fight against bacterial resistance, offering tangible benefits both at the individual level and in the healthcare system as a whole.

To conclude, this work provides a solid foundation for future clinical applications of artificial intelligence in microbiology, highlighting not only its technical feasibility, but also its strategic relevance in the face of one of the most pressing problems in contemporary medicine.

References

Bravo-Ortiz, M. A., Mercado-Ruiz, E., Villa-Pulgarin, J. P., Hormaza-Cardona, C. A., Quiñones-Arredondo, S., Arteaga-Arteaga, H. B., Orozco-Arias, S., Cardona-Morales, O., & Tabares-Soto, R. (2024). CVTStego-Net: A convolutional vision transformer architecture for spatial image steganalysis. *Journal of Information Security and Applications*, *81*, 103695. https://doi.org/10.1016/J.JISA.2023.103695.

Condorelli, C., Nicitra, E., Musso, N., Bongiorno, D., Stefani, S., Gambuzza, L. V., Carchiolo, V., & Frasca, M. (2024). Prediction of antimicrobial resistance of Klebsiella pneumoniae from genomic data through machine learning. *PLOS ONE*, *19*(9), e0309333. https://doi.org/10.1371/JOURNAL.PONE.0309333.

Dash, S., Sethy, P. K., & Behera, S. K. (2023). Cervical Transformation Zone Segmentation and Classification based on Improved Inception-ResNet-V2 Using Colposcopy Images. *Cancer Informatics*, *22.* https://doi.org/10.1177/11769351231161477/ASSET/6C0270A3-A20C-4175-8891-97D16D4C2E25/ASSETS/IMAGES/LARGE/10.1177_11769351231161477-FIG7.JPG

Gato, E., Arroyo, M. J., Méndez, G., Candela, A., Rodiño-Janeiro, B. K., Fernández, J., Rodríguez-Sánchez, B., Mancera, L., Arca-Suárez, J., Beceiro, A., Bou, G., & Oviaño, M. (2023). Direct Detection of Carbapenemase-Producing Klebsiella pneumoniae by MALDI-TOF Analysis of Full Spectra Applying Machine Learning. *Journal of Clinical Microbiology*, *61*(6), e0175122. https://doi.org/10.1128/JCM.01751-22. https://doi.org/10.1128/JCM.01751-22

Gato, E., Constanso, I. P., Candela, A., Galán, F., Rodiño-Janeiro, B. K., Arroyo, M. J., Méndez, G., Mancera, L., Alioto, T., Gut, M., Gut, I., Álvarez-Tejado, M., Rodríguez-Sánchez, B., Bou, G., & Oviaño, M. (2021). An Improved Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry Data Analysis Pipeline for the Identification of Carbapenemase-Producing Klebsiella pneumoniae. *Journal of Clinical Microbiology*, *59*(7). https://doi.org/10.1128/JCM.00800-21/SUPPL_FILE/JCM.00800-21-S0001.PDF

GeeksforGeeks (2023, May 18). *ML | Additional tree classifier for feature selection*. https://www.geeksforgeeks.org/machine-learning/ml-extra-tree-classifier-for-feature-selection/.

GeeksforGeeks. (2024, May 20). *Linear and quadratic discriminant analysis with Sklearn.* https://www.geeksforgeeks.org/machine-learning/linear-and-quadratic-discriminant-analysis-using-sklearn/

GeeksforGeeks. (2025a, January 16). *Decision tree.* https://www.geeksforgeeks.org/machine-learning/decision-tree/

GeeksforGeeks. (2025b, April 28). *Binary classification with LightGBM*. https://www.geeksforgeeks.org/machine-learning/binary-classification-using-lightgbm/.

GeeksforGeeks. (2025c, May 14). *K-Nearest Neighbors (KNN) Algorithm*. https://www.geeksforgeeks.org/k-nearest-neighbours/?utm_source=chatgpt.com.

GeeksforGeeks. (2025d, May 14). *Impulse in machine learning | Impulse and AdaBoost.* https://www.geeksforgeeks.org/machine-learning/boosting-in-machine-learning-boosting-and-adaboost/

GeeksforGeeks. (2025e, May 28). *Support Vector Machine (SVM) algorithm*. https://www.geeksforgeeks.org/support-vector-machine-algorithm/.

GeeksforGeeks. (2025f, May 30). *Multinomial naive Bayes*. https://www.geeksforgeeks.org/machine-learning/multinomial-naive-bayes/.

GeeksforGeeks. (2025g, May 30). *Random forest classifier with Scikit-learn.* https://www.geeksforgeeks.org/random-forest-classifier-using-scikit-learn/

GeeksforGeeks. (2025h, May 30). *Naive Bayesian Gaussian*. https://www.geeksforgeeks.org/gaussian-naive-bayes/?utm_source=chatgpt.com.

GeeksforGeeks. (2025i, May 30). *Gradient boosting in machine learning.* https://www.geeksforgeeks.org/ml-gradient-boosting/

GeeksforGeeks. (2025j, June 3). *Logistic regression in machine learning.* https://www.geeksforgeeks.org/understanding-logistic-regression/?utm_source=chatgpt.com

Gold, S. and R. (1996). Softmax to softassign: Neural network algorithms for combinatorial optimization. *Journal of Artificial Neural Networks*, 381-399. https://www.cise.ufl.edu/~anand/pdf/jannsub.pdf

Holguin-Garcia, S. A., Guevara-Navarro, E., Daza-Chica, A. E., Patiño-Claro, M. A., Arteaga-Arteaga, H. B., Ruz, G. A., Tabares-Soto, R., & Bravo-Ortiz, M. A. (2024). A comparative study of CNN-capsule-net, CNN-transformer encoder, and Traditional machine learning algorithms to classify epileptic seizure. *BMC Medical Informatics and Decision Making*, *24*(1), 1-23. https://doi.org/10.1186/S12911-024-02460-Z/FIGURES/11.

Hsiao, T. Y., Chang, Y. C., Chou, H. H., & Chiu, C. Te. (2019). Filter-based deep-compression with global average pooling for convolutional networks. *Journal of Systems Architecture*, *95*, 9-18. https://doi.org/10.1016/J.SYSARC.2019.02.008.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). *Densely Connected Convolutional Networks*. https://github.com/liuzhuang13/DenseNet.

Huang, T. S. S., Lee, S. S. J., Lee, C. C., & Chang, F. C. (2020). Detection of carbapenem-resistant Klebsiella pneumoniae on the basis of matrix-assisted laser desorption ionization time-of-flight mass spectrometry by using supervised machine learning approach. *PloS One*, *15*(2). https://doi.org/10.1371/JOURNAL.PONE.0228459. https://doi.org/10.1371/JOURNAL.PONE.0228459

Kong, P. H., Chiang, C. H., Lin, T. C., Kuo, S. C., Li, C. F., Hsiung, C. A., Shiue, Y. L., Chiou, H. Y., Wu, L. C., & Tsou, H. H. (2022). Discrimination of Methicillin-Resistant Staphylococcus aureus by MALDI-TOF Mass Spectrometry with Machine Learning Techniques in Patients with Staphylococcus aureus Bacteremia. *Pathogens*, *11*(5). https://doi.org/10.3390/PATHOGENS11050586/S1

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., Xie, S., & Research, F. A. (2022). *A ConvNet for the 2020s*. Retrieved May 14, 2025, from https://github.com/facebookresearch/ConvNeXt

López-Cortés, X. A., Manríquez-Troncoso, J. M., Hernández-García, R., & Peralta, D. (2024). MSDeepAMR: antimicrobial resistance prediction based on deep neural networks and transfer learning. *Frontiers in Microbiology*, *15.* https://doi.org/10.3389/FMICB.2024.1361795/PDF. https://doi.org/10.3389/FMICB.2024.1361795/PDF

Luo, P., Wang, X., Shao, W., & Peng, Z. (2018). *TOWARDS UNDERSTANDING REGULARIZATION IN BATCH NORMALIZATION*.

Simonyan, K., & Zisserman, A. (2015). *VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION.* http://www.robots.ox.ac.uk/

Tabares-Soto, R., Arteaga-Arteaga, H. B., Mora-Rubio, A., Bravo-Ortíz, M. A., Arias-Garzón, D., Alzate-Grisales, J. A., Orozco-Arias, S., Isaza, G., & Ramos-Pollán, R. (2021). Sensitivity of deep learning applied to spatial image steganalysis. *PeerJ Computer Science*, *7*, 1-27. https://doi.org/10.7717/PEERJ-CS.616/TABLE-8.

Tan, M., & Le, Q. V. (2020). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*.

Tan, M., & Le, Q. V. (2021). *EfficientNetV2: Smaller Models and Faster Training*. https://github.com/google/.

Tharmakulasingam, M., Wang, W., Kerby, M., Ragione, R. La, & Fernando, A. (2023). TransAMR: An Interpretable Transformer Model for Accurate Prediction of Antimicrobial Resistance Using Antibiotic Administration Data. *IEEE Access*, *11*, 75337-75350. https://doi.org/10.1109/ACCESS.2023.3296221.

Xu, J., Li, Z., Du, B., Zhang, M., & Liu, J. (2020). Reluplex made more practical: Leaky ReLU. *2020 IEEE Symposium on Computers and Communications (ISCC)*, *2020-July*, 1-7. https://doi.org/10.1109/ISCC50000.2020.9219587. https://doi.org/10.1109/ISCC50000.2020.9219587

Xu, X. (2024). Modelling the rapid detection of Carbapenemase-resistant Klebsiella pneumoniae based on machine learning and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Diagnostic Microbiology and Infectious Disease*, *110*(2), 116467. https://doi.org/10.1016/J.DIAGMICROBIO.2024.116467. https://doi.org/10.1016/J.DIAGMICROBIO.2024.116467

Yu, J., Lin, Y. T., Chen, W. C., Tseng, K. H., Lin, H. H., Tien, N., Cho, C. F., Huang, J. Y., Liang, S. J., Ho, L. C., Hsieh, Y. W., Hsu, K. C., Ho, M. W., Hsueh, P. R., & Cho, D. Y. (2023). Direct prediction of carbapenem-resistant, carbapenemase-producing, and

colistin-resistant Klebsiella pneumoniae isolates from routine MALDI-TOF mass spectra using machine learning and outcome evaluation. *International Journal of Antimicrobial Agents*, *61*(6), 106799. https://doi.org/10.1016/J.IJANTIMICAG.2023.106799. https://doi.org/10.1016/J.IJANTIMICAG.2023.106799

Zeng, Y., Wang, C., Ye, Q., Liu, G., Zhang, L., Wan, J., & Zhu, Y. (2023a). Machine learning model of imipenem-resistant Klebsiella pneumoniae based on MALDI-TOF-MS platform: An observational study. *Health Science Reports*, *6*(9), e1108. https://doi.org/10.1002/HSR2.1108.

Zhang, Y. M., Tsao, M. F., Chang, C. Y., Lin, K. T., Keller, J. J., & Lin, H. C. (2023a). Rapid identification of carbapenem-resistant Klebsiella pneumoniae based on matrix-assisted laser desorption ionization time-of-flight mass spectrometry and an artificial neural network model. *Journal of Biomedical Science*, *30*(1), 1-10. https://doi.org/10.1186/S12929-023-00918-2/FIGURES/5. https://doi.org/10.1186/S12929-023-00918-2/FIGURES/5.

Acknowledgments

First of all, I want to thank God, who has been my constant guide and strength every step of the way. Thank you for giving me the strength to get to this point and for allowing me to fulfill one of my biggest dreams in life: becoming a master's degree holder. On difficult days, when my strength was failing, I felt His presence urging me to continue and not give up.

To my unconditional partner, thank you for accompanying me with love, patience, and faith throughout this process. Your unwavering support, words of encouragement, and constant presence were essential in helping me not to falter and to learn to persevere even when everything seemed uphill.

To my parents, thank you for being my role models. For always believing in me, for highlighting my strengths, and for being there unconditionally. Your love and trust were the driving force that pushed me to give my best.

To my friends, thank you for being an invaluable support network. Thank you for believing in me, for encouraging me in moments of doubt, and for making me feel like a source of inspiration. Your words of admiration reminded me many times why it was worth continuing.

To Synlab, my deepest gratitude for allowing me to work with the data that made this research possible. Thank you for your willingness, your openness, and for facilitating each of the academic requirements that this work demanded. Your support was essential to bringing this project to fruition.

To Universidad EAFIT, thank you for welcoming me with such humanity and for offering me an academic experience of unparalleled quality. I am grateful for the knowledge I have gained and for the exceptional professors who became my guides along the way. You taught me not only to think critically but also to grow as a professional and as a human being.

Finally, to my thesis advisors, thank you for your constant support and for always pushing me to become a better professional and researcher. Your guidance, your patience, and your belief in my potential were instrumental to the success of this work. I am grateful not only for your academic excellence, but also for your kindness and integrity as individuals.

And lastly, I thank myself. Because without a doubt, this has been the most difficult challenge I have faced so far. I am deeply proud of this beautiful work, to which I have devoted years of effort, sacrifice, and learning. This thesis is not only an academic achievement, but an expression of my personal growth, my passion for knowledge, and my resilience.

With infinite gratitude,

Valentina Salazar Marín

COVER LETTER

June 13, 2025

To: Institute of Electrical and Electronics Engineers (IEEE)

Editors-in-Chief

IEEE Access

From: Geysson Javier Fernandez García, Mario Alejandro Bravo García, Valentina Salazar Marín*

*Corresponding autor: vsalazarm@eafit.edu.co

Dear Editors-in-chief and members of the editorial board

I am pleased to submit the manuscript entitled "**Detection of carbapenem resistance in *Klebsiella pneumoniae* using vision transformers and MALDI-TOF proteomic profiles**" for consideration for publication in IEEE Access.

This article describes a novel approach that combines MALDI-TOF mass spectrometry with a hybrid artificial intelligence architecture based on *Convolutional Vision Transformers (CVT)* to classify carbapenem-resistant and carbapenem-susceptible strains of *Klebsiella pneumoniae*. Unlike previous studies focusing on classical algorithms or high-cost genomic models, this research uses proteomic spectra treated as images, which allows for the capture of spatial relationships and complex nonlinear patterns with greater accuracy and diagnostic speed.

Among the most relevant results, the CVT model achieved an average accuracy of 80.19%, with a recall of 88.5% and an F1-score of 89% for the resistant class, far surpassing classic models such as SVM or Random Forest applied to numerical data. In addition, Grad-CAM was used to provide interpretability to the model by visualizing the spectral regions most decisive in the prediction. This represents a methodological advance with high potential for scalability and clinical impact, especially in hospital settings.

We believe that this work fits within the multidisciplinary approach of IEEE Access, integrating clinical microbiology, data science, deep learning, and emerging technologies applied to the healthcare sector. Furthermore, its innovative and clinically relevant approach makes it of interest to both the research community and professionals working in artificial intelligence applied to biomedicine.

We affirm that this manuscript is original, has not been published, and is not under review in another journal. All authors have read and approved the content, and there are no conflicts of interest to declare.

Thank you in advance for your consideration. I look forward to hearing from you with any additional comments or requests.

Sincerely,

Geysson Javier Fernandez García, Mario Alejandro Bravo García, Valentina Salazar Marín*
*Corresponding autor: vsalazarm@eafit.edu.co