

The background of the entire page is a grayscale scanning electron microscope (SEM) image showing a dense field of irregular, textured dust particles. The particles vary in size and shape, with some appearing as small, rounded grains and others as larger, more complex, interconnected structures. The lighting creates highlights and shadows, emphasizing the three-dimensional nature of the particles.

SEMAI Particle Analyzer

Morphochemical classification of
dust particles by microgrph anal-
isys

David Restrepo Rivera

Universidad EAFIT

UNIVERSIDAD
EAFIT[®]

Inspira Crea Transforma

Contents

1	Introduction	1
1.1	Abstract	1
1.2	Results	2
2	Fisical principle of the sampling technique	5
2.1	Microscopy	5
2.2	Electronic Microscopy..	5
2.3	SEM	5
2.3.1	Back scattered electrons.	6
2.3.2	Secondary electrons.	6
2.4	Energy Dispersive Spectroscopy (EDS)	6
2.5	Sample preparation.	7
3	Particle material in Airborn dust.	9
3.1	Air borne dust.	9
3.2	Characterization of particles..	9
3.3	Particle Matter Collection in Medellin, Colombia	11
3.4	ME22 Campaign Analysis and Dataset.	12
3.4.1	ME22 particle Dataset.	13
3.4.2	ME22 particle analisis.	15
3.5	Origin Based Morphochemical groups.	18
3.5.1	Organic:	18
3.5.2	Minerals:..	19
3.5.3	Metals:	19
3.5.4	Composite:..	20
4	Standard Practice for particulate material Analysis by SEM-EDS.	23
4.1	Scope	23
4.2	Terminology	23
4.3	Summary of Practice.	23
4.4	Significance and Use.	23
4.5	Sample Preparation	24
4.6	Sample Area.	24
4.7	Instrument Requirements and Operation	24
4.7.1	General.	24
4.7.2	Scanning Electron Microscope (SEM)	25
4.7.3	Energy Dispersive X-ray Spectrometry (EDS)	25
4.7.4	Sample Placement	25
4.7.5	Sample Processing and Data Acquisition. Detection and Calibration	25
4.8	Quality Control	26
5	Particle detection.	27
5.1	Particle detection by binary threshold.	27
5.2	High frequency noise removal.	28
5.3	Morphological operations on segmentation mask.	29
5.4	Connected region detection.	30
5.5	Single particle cropping..	30
5.6	Code flow	31

6 Automatic Particle Detection.	33
6.1 Review of AI Techniques in SEM Image Segmentation	33
6.2 Implementation of an automatic particle segmentation pipeline.. . . .	34
6.3 Conclusions on AI and SEM Image Segmentation	36
7 Vision particle classifier.	39
7.1 Training vision model	39
7.2 Model metrics and validation.	40
7.3 Model analysis and conclusions.	41
8 Semparticle Analyzer: A Web Application for SEM Micrograph Analysis	43
Bibliography	45

1

Introduction

1.1. Abstract

The central objective of this study is to meticulously classify dust particles through a detailed microscopic analysis. To achieve this, the specific goals include a comprehensive characterization of particles, the development of a robust methodology for morphochemical classification, and the generation of an extensive dataset, comprising over 140 descriptors, to facilitate in-depth investigations.

The outcomes of this work are significant. Successful detection and analysis of particles, utilizing scanning electron microscopy (SEM) micrographs and energy-dispersive X-ray spectroscopy (EDS), contribute to a comprehensive dataset. This dataset not only quantifies particulate matter characteristics but also offers nuanced insights into its composition. This information has the potential to unravel intricate environmental implications associated with particulate matter, elevating its significance in broader ecological contexts.

The physical principles involved in SEM (scanning electron microscopy) and EDX (energy-dispersive X-ray spectroscopy) are described in { include the interaction of the electron beam with the specimen, emission of secondary electrons and backscattered electrons, and the analysis of characteristic X-rays emitted from the material when it is ionized by an incident electron. SEM provides information on the topography, morphology, composition, and crystallographic nature of the analyzed specimens. The emission depth of the different signals used in SEM is influenced by the electron beam energy, specimen nature, composition, and sample preparation. In the case of EDX, the process involves the acquisition and analysis of characteristic X-rays emitted from a material when it is ionized by an incident electron. This signal is essential for analytical purposes and provides information on the qualitative and quantitative analysis of the specimens.

Airborne dust particles in the study are then characterized using a combination of scanning electron microscopy (SEM), energy-dispersive X-ray spectroscopy (EDX), and machine learning (ML) classification. This methodology allows for the identification and quantification of various particle groups present in the airborne dust samples. The selected groups for characterization include:

The morphochemical groups identified in the analysis are organic particles, mineral particles, metallic particles, tire wear particles, and salt particles. These groups were selected based on their distinct characteristics, as revealed by the chemical composition and elemental constituents of the particulate matter samples.

Organic Particles:

Predominantly composed of carbon (C) with varying concentrations (50Presence of oxygen (O).

Mineral Particles:

Primarily composed of silicon-oxygen (SiO) crystals. Smaller quantities of calcium (Ca) and aluminum (Al) were also detected. Metallic Particles:

Consistently identified as iron (Fe), often with varying degrees of oxidation. Tire Wear Particles:

Exhibit composite compositions with different ratios of carbon (C), oxygen (O), iron (Fe), aluminum (Al), silicon (Si), and calcium (Ca). Salt Particles:

Distinct composition containing sodium (Na) and chlorine (Cl). These characteristics provide valuable insights into the composition and elemental constituents of each morphochemical group within

the analyzed particulate matter samples.

These groups were chosen to provide comprehensive insights into the composition and sources of the particulate matter, aiding in the understanding of potential environmental impacts and sources of pollution.

The standard practice for particle analysis involves the use of scanning electron microscopy/energy-dispersive X-ray spectrometry (SEM/EDS) to analyze airborne particulate material. The analysis is performed manually to screen the sample for candidate particles associated with specific morphochemical groups. The practice includes the following key aspects:

Quality Control:

Establishing a protocol to confirm appropriate instrument operation parameters prior to sample analysis. Monitoring the EDS X-ray energy calibration and SEM beam current stability using appropriate reference materials. Analyzing a reference material with particles of known size range and composition to test the accuracy of particle detection and identification. Incorporating environmental or control samples into the analytical protocol to monitor the cleanliness of the sample collection or analytical system. **Sample Processing and Data Acquisition:**

Visual differentiation of particles at x500 magnification to identify characteristic morphochemical groups. Recording the positions of the stubs on the SEM stage when the samples are inserted. Conducting particle imaging and capturing EDS spectra for particles of special interest. **Non-Destructive Analysis:**

Unlike bulk sample methods, particle analysis preserves the size, shape, and individual particle characteristics, enabling the detection and identification of different morphological groups within the particulate material without dissolving or extracting it. **Sample Preparation:**

Careful handling of the sample to avoid contamination. Placing the sample substrate in a specimen stub suitable for high vacuum microscopy. Coating the sample substrate to increase its electrical conductivity, unless using an environmental SEM or variable pressure/low-vacuum SEM. These practices are aimed at ensuring accurate and reliable analysis of particulate material using SEM/EDS.

The methods used for particle detection over the SEM images included the application of binary thresholding and the utilization of artificial intelligence (AI) techniques such as deep learning and machine learning algorithms, particularly convolutional neural networks (CNNs). Binary thresholding was employed to separate particles from the background by setting an appropriate threshold value based on the intensity of the particle regions. On the other hand, AI techniques, especially CNNs, were trained on annotated SEM images to achieve high accuracy in segmenting various features like cracks, pores, and grain boundaries. The integration of AI techniques with traditional DIP methods also showed promising results in enhancing the accuracy and efficiency of SEM image segmentation.

The results of these methods were promising. The AI-based method utilizing computer vision and machine learning demonstrated significantly improved performance in accuracy, precision, recall, and f1-score compared to a baseline approach. The study also showed a strong correlation between manually measured sizes and algorithm-generated sizes, with Spearman coefficients of 0.91 for the fully automated process and 0.99 for the semi-automated process. Additionally, the application of CNNs and modified weight functions in analyzing SEM images of nanomaterials and geological features exhibited effectiveness in discriminating between complex shapes and distinguishing between different geological features. In the specific context of Medellín, Colombia, the study undertakes a meticulous examination of PM composition and levels, aiming to contribute to a nuanced understanding of air pollution in the Valle de Aburrá region. The inquiry into types, quantities, and characteristics of particles at various stations reveals a tapestry of information. Mineral particles, with varying sizes, exhibit distinct concentrations across different stations, mirroring the spatial complexity of particulate matter distribution. Additionally, the prevalence of tire wear particles and metal oxides exhibits station-specific patterns, accentuating the localized nuances of air quality.

1.2. Results

These findings provide invaluable contributions to comprehending PM_{10-2.5} dynamics in the Valle de Aburrá region. The insights empower stakeholders, including policymakers, researchers, and environmentalists, with informed decision-making capabilities. This facilitates the design and implementation of targeted strategies and interventions aimed at ameliorating air quality conditions in Medellín.

Examining the conclusive aspects of the work related to dust particle classification, the outcomes

underscore the efficacy of the vision model employed. The experiment yields promising results, with an impressive accuracy of 0.88, indicative of the model's proficiency in classifying particles. Notably, the model exhibits robust performance in categorizing BiogenicOrganic and Mineral particles, achieving commendable precision, recall, and F1-scores.

Detailed examination of class-specific metrics further elucidates the model's discriminatory capabilities. The Mineral class, for instance, attains a precision of 0.89, denoting high accuracy in identification. The Tirewear class, while slightly lower in precision at 0.77, still demonstrates substantial proficiency. The macro-averaged metrics, serving as a comprehensive measure, reveal an average precision, recall, and F1-score of 0.86, 0.81, and 0.83, respectively.

Furthermore, the distribution analysis of particle counts across various groups augments the findings, unraveling crucial insights into the composition and sources of particulate matter. The distinct categories, such as Metallic, Mineral, Salt, Ti-rich Paint, and Tire wear, collectively contribute to a nuanced understanding of environmental impacts and pollution sources. This holistic perspective substantiates the consequential implications of the research outcomes in advancing our comprehension of particulate matter dynamics.

2

Fisical principe of the sampling technique

2.1. Microscopy

Microscopy is an imaging technique that transforms objects outside the range of the unaided eye resolution into much larger images. For centuries, optical-light microscopy (OM) was the only imaging system commonly used in science. It is a simple system in which the object of interests is illuminated by a transmitted light source such as a halogen lamp, lasers, or LEDs. A condenser lens and densifies the light into a more intense one that latter reaches the specimen fixed onto a mechanical stage that enables movement under a revolving nosepiece which holds several lenses, the objective lenses with different augmentations. The objective lens collects the light, and a focused real image is formed, finally the real image is again enlarged by the ocular lenses giving the final magnified image of the observed specimen. OM has several limitations, and the most important one is the loss of resolution at very high magnifications. By applying different physical principles in image generation, several other categories of microscopy have been developed to overcome those restrictions, one of the most important, electron microscopy (EM), is widely used in materials research as well as in life sciences.

2.2. Electronic Microscopy.

The main modification of the electron microscopy is the "illumination" source which in EM is a highly energetic electron beam. The concept is easy to understand if we remember that light as radiation has a wavelength of around 400 700 nm allowing magnifications between 500x and 1500x providing visualization of structures in the scale of micrometers, while the electrons wavelengths vary between 0.001 and 0.01 nm enabling a theoretical resolution of around 0.02 nm, for a 100 kV electrons beam.

One of the main motivations in the improvement of EM is to straightforwardly attain subatomic imaging resolution and elemental analysis.

Nowadays, the occurrence of microplastics in the environment needs to be properly addressed as well as their ecological effects after their release into aquatic systems, terrestrial media, and atmosphere. The characterization of such microparticles progressed over the past decade, and optical microscopy was one of the first tools used for visual inspection and determination of their shape and size. The amount of identifiable particles by OM is restricted due to its spatial resolution, whereas such undetected particles can easily be observed under the electron microscope, enabling accurate size distribution and morphology evaluation. OM and EM are used for very different purposes.

2.3. SEM

In SEM, a highly energetic and focused electron beam is used to scans the specimen in a raster scan pattern. Then, different effects can result from the interaction between the electron beam and the atoms in the sample. Although most of the electron beam energy ends up in the specimen as heat, other events take place and are detected outside the specimen, these subsequent effects can include the emission of backscattered electrons, secondary electrons, and characteristic X-rays.

Part of the electron beam will be unscattered, though most of the electron beam interacts with the specimen and undergoes inelastic and elastic collisions. In the first case, the direction of the primary electrons is changed but their overall energy is kept. Inelastic scattered electrons change their direction and lose part of their energy.

2.3.1. Back scattered electrons.

Back-scattered electrons are the result of elastic collisions between an incident electron and an atom in the specimen, which results in a change in the electrons' trajectory losing part of its energy and being scattered 'backwards'. This type of electron originates from a broad region within the interaction volume and usually come from deeper regions of the sample than the SE. The production of backscattered electrons is directly proportional with the atomic number of the elements presents in the specimen: the higher the atomic number, the brighter that region will appear. A good illustration to understand the phenomena could be imaging the electron-atom collision as the so-called "billiard-ball" model, where small particles (electrons) collide with larger particles (atoms). Larger atoms are much stronger scatterers of electrons than light atoms, and therefore produce a higher signal. Some of the backscattered electrons can generate more secondary electrons when exiting the specimen [1]

2.3.2. Secondary electrons.

Secondary electrons (SE) originate from the surface or the near-surface regions of the sample, these electrons are generated when inelastic collisions change the path of the incident electrons transferring part of its energy to electrons in the atom of the specimen causing it to leave the sample with a very small energy. Each primary or incident electron can produce several secondary electrons, thus secondary electrons are abundant and very useful for inspecting the topography of the sample's surface and are the most used imaging signal in SEM.

A key factor in SEM is the interaction volume, the region where the electron beam penetrates the specimen, resulting in a three-dimensional volume where the electron beam interacts with the atoms of the specimen. The emission depth of the different signals used in SEM is influenced by the electron beam energy, specimen nature, composition, and sample preparation. The higher the accelerating voltage applied on the electron beam, the larger the interaction volume. In SEM, imaging is carried out by collecting the emission of secondary electrons (topography) and backscattered electrons (atomic number). Analytical X-rays enable quantitative analysis of the chemistry in the specimens. Thus, SEM provides information on the topography, morphology, composition, and crystallographic nature of the analyzed specimens.

2.4. Energy Dispersive Spectroscopy (EDS)

X-rays emission results from the deenergization of an atom in the sample after a secondary electron is produced. This signal is extremely important for analytical purposes, and it will be addressed in this section. [2]

Energy Dispersive Spectroscopy (EDS) is the process for the acquisition and analysis of the characteristics X-rays that are emitted from a material when it is ionized by an incident electron.

Back-scattered electrons result from the collision between an incident electron and an atom in the specimen and losing part of its energy and being scattered 'backwards' 180 degrees. Some of the backscattered electrons can generate more secondary electrons when exiting the specimen. The production of backscattered electrons varies directly with the atomic number of the chemical elements present in the specimen: the higher the atomic number, the brighter that region will appear. X-rays emission results from the deenergization of an atom in the sample after a secondary electron is produced. This signal is extremely important for analytical purposes, as it will be addressed in the next section. A key factor in SEM is the interaction volume, the region into which the electron beam penetrates the specimen, i.e., the three-dimensional volume resulting from the interaction of the electron beam with the specimen atoms. The emission depth of the different signals used in SEM is influenced by the electron beam energy, specimen nature, composition and sample preparation. The higher the accelerating voltage applied on the electron beam, the larger the interaction volume. In SEM, imaging is carried out by using the emission of secondary electrons (topography) and backscattered electrons (atomic number). Analytical X-rays enable qualitative and quantitative analysis of the specimens. Thus, SEM provides information on the topography, morphology, composition and crystallographic nature of

the analysed specimens.

X-rays are highly energetic photons resulting from electronic transitions in the atoms when targeted by an incident accelerated electron beam. Typically, characteristic X-rays are formed when an incident electron hits an atom of the specimen causing inner shell ionization ejecting an electron from the inner shells (K, L or M) and creating a vacancy or hole that is left in that shell, when an electron from another shell fills in that vacancy (electron transitions), X-rays are emitted, the energy of the emitted X-rays is equal to the differences in ionization energies of the electrons within the transitions. Electronic transitions to the K-shell are named KX-rays, those to the L-shell are LX-rays and to the M-shell are the MX-rays. These transitions are characteristic of each chemical element, and this is the main reason that led to the development of EDS detection systems in electron microscopy and consequently widening its applications in microstructural characterization.

The collected characteristic X-rays are the analytical signals used in electron microscopy for chemical analysis. An X-ray spectrum emitted by the specimen provides both qualitative and quantitative information, allowing identification of which elements are present in the sample and the amount of each element.

As mentioned earlier, the interaction volume from which the X-rays are generated are key factors for the quality of the final EDS spectrum. Moreover, the amount of generated signal strongly depends on the energy of the X-rays and the average atomic weight of the sample. For example, X-rays such as carbon Ka are easily absorbed by the solid sample and only a few are detected. In contrast, harder X-rays such as iron Ka are able to rupture the solid specimen and only a small amount is absorbed.

Furthermore, increasing the spatial resolution of the X-ray microanalysis could also compromise the detection for a particular chemical element, the spatial resolution is the smallest distance between two volumes from which independent X-ray microanalyses are obtained. Since the interaction volume depends on the incident beam diameter as well as the spreading of the beam caused by elastic scattering within the sample, the microanalysis spatial resolution will strongly depend on the specimen. In other words, the higher the spatial resolution, the smaller is the analyzed volume and, consequently, the smaller is the signal intensity for a particular element.

2.5. Sample preparation.

The preparation of specimens for examination in the SEM is the most important step for an effective imaging acquisition and accurate analysis. The electron beam is highly energetic, and the number of electrons hitting the specimen is higher than the ones leaving. Accordingly, there is an excess of electrons staking on the surface and generating charges, so it is compulsory that the sample is electrically conductive. Moreover, the surplus of electrons needs to be grounded otherwise incoming primary electrons are repelled by the negatively charged surface and the image becomes distorted by the charging effects. In the case of nonconductive samples such as ceramic materials, polymers or biological specimens, it is necessary to cover the surface of the sample with a very thin layer of a conductive material, for example, carbon by means of high vacuum evaporation coating or gold, gold/palladium alloy or platinum thin films using a plasma sputter coater.

Surface coating by vacuum is based on heating a conductive material to its vaporization temperature, followed by condensation of the evaporated atoms onto the surface of the sample, under high vacuum. In a typical plasma sputter coater, the conductive material is the cathode, and the sample is placed on the anode electrode. The applied voltage between the electrodes produces plasma directed by a magnetron on the sample surface. In this way, the negatively charged conductive metal surface (placed on the cathode) is sputtered by the ionized gas and the sputtered metal coats the sample surface.

In both methods the thickness of the conductive films is usually around 10 nm, depending on the nature of the sample and its conductivity requirements. The film should not be too thick since it may mask the fine details of the surface or interfere with the other signals that may be detected such as X-ray emissions. For example, if carbon is one of the elements of interest to be identified or eventually semi quantified, one should use a different conductive material to not interfere with the EDS analysis. On the other hand, a too thin film may not be enough to prevent degradation by beam heating, radiation damage or specimen volatility. Since the EM is kept in high vacuum to avoid electron scattering and produce the electron beam as monochromatic as possible, the effect of specimen volatility is minimized to a certain point [5,6,8].

There are many methods and several techniques that are applied in sample preparation for EM

observation and analysis. Nonetheless, one should keep in mind that the best sample preparation method is most probably the less invasive one and the best sample is always the one that remains closest to its natural state

An specific sample preparation method for the specific case of airborne dust material which allegedly contains organic and metallic particles detailed in Chapter 5.

3

Particle material in Airborn dust.

3.1. Air borne dust.

Particles in the atmosphere are the mixture of solid particles, liquid droplets, and liquid components in solid particles. According to their sources, particles can vary in concentration, as well as in their physicochemical and morphological characteristics. Particles can be a product of combustion, suspension of soil materials, and sea spray and can also be formed from chemical reactions in the atmosphere. It can be concluded that airborne particulate matter (PM) is a complex mixture of many different chemical species originating from a variety of sources [3]

These particles are conditioned to the aerodynamic diameter and thus classified as coarse (2.5–10 μm), fine (0.1–2.5 μm), and ultrafine ($<0.1 \mu\text{m}$), where the degree of toxicity becomes greater for smaller particles. These particles can get into the lungs and translocate into vital organs due to their size, causing significant human health consequences. Besides, PM pollutants have been linked to respiratory conditions, genotoxic, mutagenic, and carcinogenic activity in human beings.

Up to this date, most of the epidemiological studies researched particulate matter in relation to the size of particles, as it seemed that is the most influential health factors; the smaller the particles, the more evident the health effects. However, other particle characteristics such as shape, chemical composition and density also affect the particle deposition in the respiratory system, hence they also have an important role in those health effects [4]. Furthermore, the exposure to pollutant from traffic and other combustion sources have been proven to place an impact on human health more severely than other sources of PM, for example, natural, implying that the health effects associated with PM exposition would be related mostly to anthropogenic emission sources [5]. Whether the health problems are mainly related to PM physical characteristics (namely size, shape and mass) or to chemical composition are still uncertain, some studies suggest that health effects may be related more to PM specific components rather than mass, showing the importance of PM chemical characterization.

To determine size, shape and elemental compositions of PM, a wide range of techniques have been employed. Nevertheless, those techniques do not provide information concerning the composition of particles. Such information can be obtained through Scanning electron microscopy (SEM) coupled with a energy dispersive spectrometer that has proved to be highly efficient method for chemical characterization. The applications of individual particle analysis not only reveal the morphological and physical characteristics of aerosol particles but also provide evidences about individual sources, making the elemental composition of single atmospheric particles sometimes more useful than their bulk elemental composition with view to establish their origin and their potential effects on human health.

3.2. Characterization of particles.

A study conducted in Agra [6], India used SEM-EDS analysis to examine airborne particles (PM_{2.5} and PM₁₀) and identify their sources based on chemical composition and morphology. Sampling took place at a roadside site affected by traffic emissions and a semi-rural site. At the roadside, PM_{2.5} particles were spherical, cluster, flaky, and chain-like, mainly composed of carbon and oxygen-rich particles indicating soot and tarballs from vehicle exhaust. Trace amounts of sulfur and potassium were also found.

In the semi-rural area, PM_{2.5} particles had flaky, cluster, and irregular shapes, primarily consisting of oxygen and carbon-rich particles with additional elements like silicon, aluminum, magnesium, sodium, iron, calcium, and potassium, suggesting crustal and geological origins. Comparing both sites, carbon dominance was observed at the roadside, while oxygen was dominant in the semi-rural area. Elements like silicon, aluminum, sodium, sulfur, potassium, and iron were present in both areas but with higher percentages in the semi-rural site. Magnesium and calcium-rich particles indicated probable sources such as fertilizer use, vegetative burning, ceramics, refractory brick plants, and coal burning. Correlation analysis showed strong relationships between various elements and particulate matter, suggesting similar sources. Carbon correlated with aluminum, potassium, and iron, while sulfur correlated with oxygen, implying a connection to fuel-oil combustion. Silicates, likely from crustal sources, were prevalent in the semi-rural site's PM_{2.5}. In conclusion, the study emphasized the disparities in particle shape, size, and morphology between PM_{2.5} and PM₁₀ at the roadside and semi-rural sites. Chemical analysis identified carbon-rich particles at the roadside and crustal elements in PM₁₀ at both sites, originating from sources like vehicular emissions, coal burning, soil dust, and natural sources. Understanding particle composition and morphology is crucial for developing effective pollution mitigation strategies.

In the article of our collaborators Raus et. al [7] presents a state-of-the-art methodology for automated identification and quantification of tire wear particles (TWP) in airborne dust using a combination of scanning electron microscopy (SEM), energy-dispersive X-ray spectroscopy (EDX), and machine learning (ML) classification. The study employed an extensive set of 67 morphological, textural, and chemical descriptors to differentiate environmental particles into classes such as TWP, metals, minerals, and biogenic/organic particles. The ML model was developed and trained using over 100,000 environmental particles, including 6,841 TWP, and applied in a one-year monitoring campaign at two sites in Switzerland. The mass concentrations of TWP in different airborne fractions were determined, and the particle size distribution and shape characteristics of TWP were evaluated. Additionally, the internal structure of TWP was characterized using focused ion beam scanning electron microscopy (FIB-SEM). The study demonstrated that TWP contained mineral and metal particles not only on their surfaces but also throughout their volumes. The results showed variations in TWP concentrations between the urban background and urban kerbside sites. The ML-based classifier achieved high accuracy in particle classification, correctly identifying about 98/

Previous work on EAFIT university [8] focuses on the characterization of coarse particulate matter (PM_{10-2.5}) in the Valle de Aburrá region of Medellín, Colombia, with the aim of understanding its sources. Samples were collected using a passive sampler (Sigma-2) with boron substrates that facilitate the analysis of carbon-containing material. The collected particles were analyzed using automated scanning electron microscopy (SEM) and energy-dispersive X-ray spectroscopy (EDX) techniques to extract topological and chemical composition information. Morpho-chemical features, such as fractal dimension, convexity, solidity, circularity, and chemical characteristics, were derived from the microscopy and spectroscopy data and used to train a classification model for identifying the source of pollution.

The objective of the study was to contribute to the characterization of PM_{10-2.5} in Medellín using automated SEM/EDX analysis of individual particles. This approach enables the morpho-chemical characterization of particles to determine their concentration, composition, size, and provide insights into their sources of origin.

The methodology involved sampling at three different stations with different typologies (urban, sub-urban, and high vehicular flow with nearby industries) in Medellín over a two-month period. Boron substrates were used for particle collection, allowing the quantification of carbon content using SEM/EDX. The collected substrates were analyzed at a laboratory using an automated SEM/EDX system. The analysis involved extracting backscattered electron (BSE) images and elemental spectra (EDX) of individual particles. The SEM/EDX data, along with morphological parameters and elemental composition, were used to classify particles into specific morpho-chemical groups associated with known emission sources.

Concentration calculations for coarse particulate matter followed the VDI2119:2013 standard, which accounts for particle mass, density, collection area, exposure time, and sedimentation velocity. The aerodynamic diameters of particles were determined, and their contributions to the PM_{10-2.5} concentration were quantified using an efficiency curve.

The results showed the concentrations, size distribution, and chemical composition of particles for

each station and period. Mineral particles were found to have the highest concentration in E1 station, with larger minerals in E2 and smaller minerals in E3. Tire wear particles were most concentrated in E2, while metal oxides had the highest concentration in E3. The study provides valuable insights into the sources and characteristics of PM_{10-2.5} in the Valle de Aburrá region, contributing to a better understanding of air pollution in Medellín, Colombia.

3.3. Particle Matter Collection in Medellin, Colombia

The Particle Matter Sampling Campaign in Medellin, Colombia aimed to gather crucial data on airborne particle matter (PM) in three distinct sampling sites. The campaign focused collecting particulate material in the PM_{2.5}-PM₁₀ range to assess PM levels and composition at each location to gain insights into the air quality dynamics within the city.

The first sampling site, Museo de Antioquia, served as an urban station located in the bustling city center. Here, a wide range of sources emitted particles of various types, making it an ideal location to study the urban PM landscape. The objective was to investigate the composition and concentration of particles emanating from different urban sources, contributing to a comprehensive understanding of the air quality in densely populated areas.

The second sampling site, Tanques la Y, was strategically positioned near a moderately busy highway situated on an inclined road at the outskirts of the city, allowed for the examination of particle matter influenced by heavy traffic emissions. The goal was to analyze the impact of vehicular exhaust and related sources on PM composition, considering the significant volume of vehicles passing through this area.

The third sampling site, Tanques Girardota, provided a contrasting environment as a rural location undergoing construction projects. This site allowed for the assessment of particle matter associated with construction activities in a less urbanized setting. The objective was to evaluate the impact of construction-related dust and particles on air quality in rural areas.

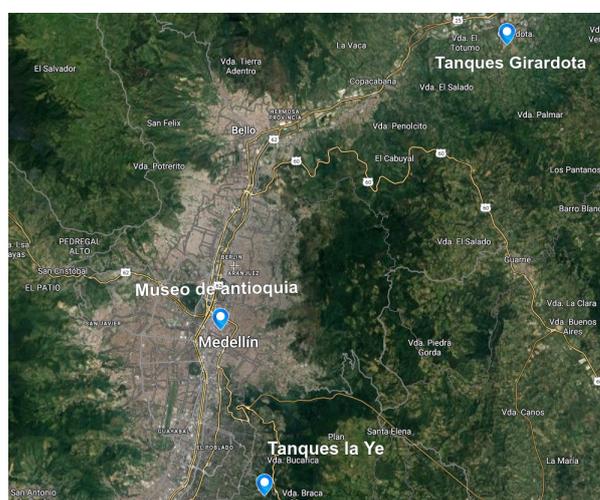


Figure 3.1: Map of sampling sites.

For the sampling campaign, the Sigma-2 passive sampler was chosen as the primary device. The Sigma-2 is a passive sampling system that does not require external power for PM₁₀ capture. The sampling duration for each site was set to a minimum of two weeks, ensuring sufficient time to collect an adequate sample representative of the respective environments.



Figure 3.2: Sigma-2 particle sampler.

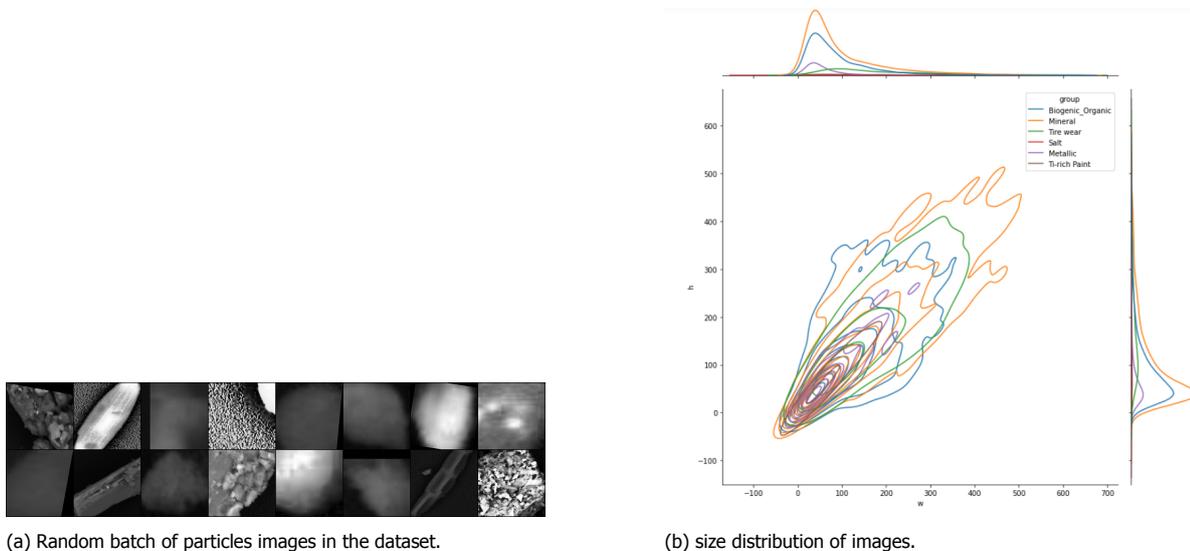
By employing this sampling methodology, the campaign aimed to gather comprehensive data on PM composition and levels at each site. The collected samples will undergo rigorous analysis to determine the types and quantities of particles present. The findings will provide valuable insights into the air quality conditions in Medellin, enabling policymakers, researchers, and environmentalists to make informed decisions regarding mitigation strategies and interventions to improve the overall air quality in the city.

After the sampling period, the collected samples were sent to the ParticleVision laboratory for analysis. ParticleVision employed the state-of-the-art methodology described earlier for the identification and quantification of tire wear particles (TWP) in airborne dust samples. This methodology uses a combination of scanning electron microscopy (SEM), energy-dispersive X-ray spectroscopy (EDX), and machine learning (ML) classification. It employed a comprehensive set of 20 morphological, textural, and chemical descriptors to differentiate environmental particles into classes such as TWP, metals, minerals, and biogenic/organic particles. This approach enabled precise identification and quantification of morphological groups within the samples. We would like to express our gratitude to Dr. XXXX XXX for his research and the development of the methodology. His contributions have been crucial for the development of this project.

3.4. ME22 Campaign Analysis and Dataset.

The preliminary results of the experiment focusing on the composition of particulate matter reveal significant findings. Following the collection of particles during the campaign, the particles were sent to the Particle Vision Laboratory for in-depth analysis. The analysis encompassed the detection of 7420 particles, each of which was subjected to analysis through DIP of (SEM) micrographs and EDS. The cropped micrographs of each particles where saved and allowed for a detailed examination of particle morphology. Moreover, the analysis yielded a comprehensive dataset consisting of over 140 descriptors, including information composition through EDS signals, concentration and size for each particles, and a big variety of morphological features. This wealth of information provides valuable insights into the composition and characteristics of the particulate matter, laying the foundation for further investigations and potential environmental implications.

3.4.1. ME22 particle Dataset.



(a) Random batch of particles images in the dataset.

(b) size distribution of images.

Figure 3.3: ME22 dataset images.

The dataset comprises a comprehensive collection of information regarding particles characteristics, it provides valuable insights into their morphology and composition. The columns can be grouped into several categories to facilitate a thorough understanding of the dataset.

1. Sample and Particle Information:

- Sample: This column contains identifiers or labels for the samples from which the particles were collected. It helps in distinguishing different sources or locations.
- Particle ID: Each particle a sample is assigned a unique identification number or code, enabling individual particle tracking and analysis.
- Group: This column categorizes particles into morphochemical groups based on the classification perform in [XXXXXXxx], for comparative analysis of particle types.
- Score: The score represents quantitative quality of the classification associated with each particle, indicating its significance or relevance within the dataset.

2. Particle Concentration and Size:

- PM80-1 ($\mu\text{g}/\text{m}^3$): This column provides the concentration of particles with sizes less than or equal to $80 \mu\text{m}$ per cubic meter of air. It helps in understanding the overall particle load in the environment.
- Particle # conc. ($\#\text{particle}/\text{m}^3$): It indicates the concentration of particles per cubic meter of air, providing information about the density of particles in the sampled area.
- dAero (μm): This column represents the aerodynamic diameter of the particles, which is a crucial parameter in assessing their behavior and potential health impacts.
- LengthMicron, WidthMicron: These columns denote the dimensions of the particles, typically measured in micrometers. They contribute to understanding the physical morphology of the particles.

3. Morphological Features:

- ECDMicron: Equivalent circular diameter of particle expressed in micrometers.
- AreaMicron²: It represents the surface area of the particles, providing insights into their size and potential interactions with the environment.

- Circularity, Compactness, Convexity: These columns measure different aspects of particle shape and morphology, such as circularity, compactness, and convexity. They help characterize the geometric properties of the particles.
- AspectRatio: This column quantifies the ratio between the length and width of the particles, aiding in understanding their elongation or distortion.

4. Elemental Composition (EDS):

- H, C, N, O, and other elements: These columns represent the presence and concentrations of various elements detected using Energy Dispersive Spectroscopy (EDS) analysis. They provide insights into the chemical composition and elemental makeup of the particles.

The dataset encompasses a wide range of parameters, from sample and particle identification to morphological features and elemental composition. It offers a comprehensive understanding of the particles' physical properties, concentrations, and composition, enabling researchers to explore relationships between particle characteristics and potential environmental implications.

	Sample	Particle Id	Group	Score	PM80-1 (µg/m3)	Particle # conc. (#particle/m3)	CmNormPM10-PM2.5 (µg/m3)	Mass deposition (µg/cm2/s)	Number deposition (#particles/cm2/s)	MassMicrogram (µg)	...	Th	Pa	I
0	BO22_Car_p1_sec_analysis	16	Biogenic_Organic	0.961	0.048785	298.552352	0.042244	6.920258e-15	0.000042	0.000163	...	0	0	
1	BO22_Car_p1_sec_analysis	22	Biogenic_Organic	0.972	0.042139	400.144760	0.038623	4.459923e-15	0.000042	0.000105	...	0	0	
2	BO22_Car_p1_sec_analysis	130	Biogenic_Organic	0.963	0.030344	771.693922	0.026766	1.665273e-15	0.000042	0.000039	...	0	0	
3	BO22_Car_p1_sec_analysis	132	Biogenic_Organic	0.963	0.023648	1270.575606	0.019901	7.882287e-16	0.000042	0.000019	...	0	0	
4	BO22_Car_p1_sec_analysis	150	Biogenic_Organic	0.780	0.021368	1556.176408	0.015646	5.815202e-16	0.000042	0.000014	...	0	0	
...
7415	ME22_TLY_P2	1446	Tire wear	0.973	0.035577	136.383237	0.023998	4.658701e-15	0.000032	0.000261	...	0	0	
7416	ME22_TLY_P2	1500	Tire wear	0.989	0.024097	297.297857	0.022019	1.447500e-15	0.000032	0.000081	...	0	0	
7417	ME22_TLY_P2	1614	Tire wear	0.384	0.029359	200.269512	0.024489	2.618087e-15	0.000032	0.000147	...	0	0	
7418	ME22_TLY_P2	1662	Tire wear	0.964	0.017879	540.049822	0.016994	5.912288e-16	0.000032	0.000033	...	0	0	
7419	ME22_TLY_P2	1701	Tire wear	0.801	0.016186	658.881086	0.015200	4.387284e-16	0.000032	0.000025	...	0	0	

420 rows × 144 columns

Figure 3.4: sample dataframe with particle features.

Among the different groups, the most predominant is the BiogenicOrganic group, accounting for 2355 particles. These particles are likely derived from organic materials of natural origin. The Metallic group consists of 465 particles, indicating the presence of metallic particles in the sampled particulate matter. The Mineral group, with 3716 particles, suggests the prevalence of particles originating from mineral sources. The presence of only 12 particles classified as Salt suggests a relatively low contribution from salt particles in the analyzed samples. Additionally, the dataset includes 83 particles categorized as Ti-rich Paint, indicating the presence of particles containing titanium-rich paint components. Lastly, the Tire wear group comprises 789 particles, suggesting the presence of particles resulting from the wear and degradation of tires. The distribution of particle counts across these groups provides insights into the composition and sources of the particulate matter, aiding in the understanding of potential environmental impacts and sources of pollution.

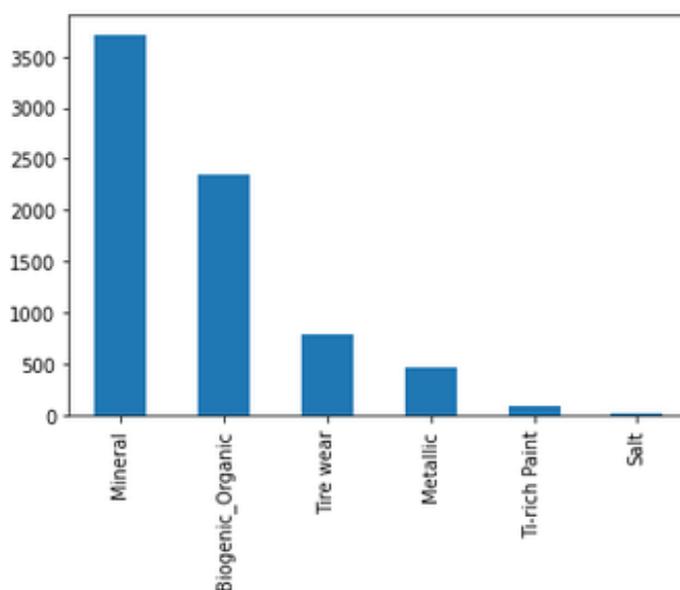


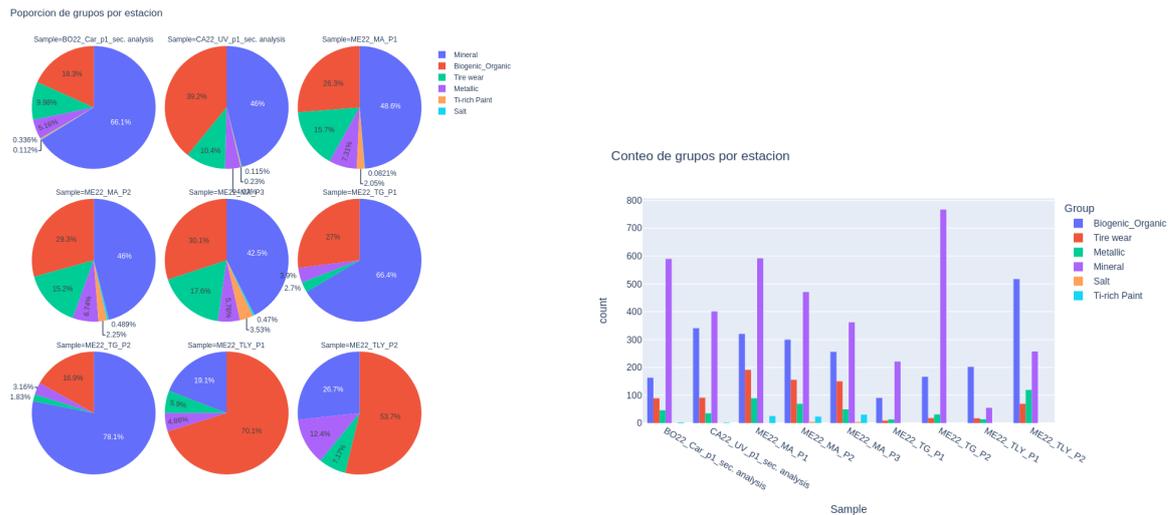
Figure 3.5: Group distribution for particle samples.

3.4.2. ME22 particle analysis.

The analysis of the database provides valuable insights into the composition and characteristics of the particulate matter.

The pie and bar distributions of groups per sample reveal interesting patterns. The predominance of groups vary across samples, indicating possible associations between specific groups and sample sources. For instance, all the samples but TY show a higher proportion of Mineral particles compared to other groups, suggesting a potential contribution from mineral materials in different places of the city. Mineral particles indicate a high amount of anthropogenic sources such as construction, industry, and resuspension of road particles. Notably, in the TG sample, a considerable amount of mineral particles were detected. This finding aligns with the location of TG in a rural area undergoing road construction works during the sampling period. In TY, the predominant contribution of BiogenicOrganic suggests a potential reduction of anthropogenic sources in the analyzed area. BiogenicOrganic particles, originating from natural sources such as vegetation and biological processes, indicate a lesser influence of human activities on particulate matter pollution. The three samples from MA, BO, and CA display a higher prevalence of TireWear composite and metallic particles, indicating a potential influence of vehicular sources in these areas. The TireWear composite is prominently present in all three samples, with the MA sample showing the highest proportion, followed by the BO station and the Y tank sample. This suggests a significant volume of vehicular traffic in these locations, leading to tire wear and associated particles generation. Metallic particles, which typically result from the suspension of brake material as the result from the abrasion between tires and brakes, also contribute to the samples composition. The distribution of metallic particles shows a similar trend across the three sites, with a peak concentration observed in the TY sample. This concentration peak could indicate a higher vehicular flow passing through the descending road near the station during the sampling period, further contributing to the elevated levels of metallic particles.

Salt and Ti Rich particles are exclusively observed in urban stations, with a higher concentration observed in the MA sample. The presence of Salt particles in urban areas can be attributed to tinsmithing work and painting as proposed by [7]. Additionally, the Ti Rich particles are commonly associated with painting supplies which often contain titanium as a component and are usually used in the context of carpentry and auto workshops, where painting is frequently performed, the release of Ti-rich particles is expected. Therefore, the higher prevalence of Ti Rich particles in the MA sample can be attributed to the significant carpentry and auto-related operations in the area. These findings emphasize the impact of localized human activities, such as carpentry and painting, on the presence of Salt and Ti Rich particles in urban environments, particularly in areas with extensive industries like MA.



(a) Concentration of groups per sample station.

(b) Count of particles per group per sample.

Figure 3.6: Group count and distribution per samples.

The size distribution plot reveals a consistent pattern across all samples characterized by a right skewed Gaussian distribution with a noticeable right tail. The mean particle equivalent circular diameter is centered around 1.5 micrometers, indicating a predominant presence of fine particles in the PM1-2.5 range. However, larger particles extending up to 8 micrometers can also be observed, although they occur less frequently. This size distribution provides insights into the overall particle size profile in the analyzed dataset.

Histogram of ECDCMicron per sample

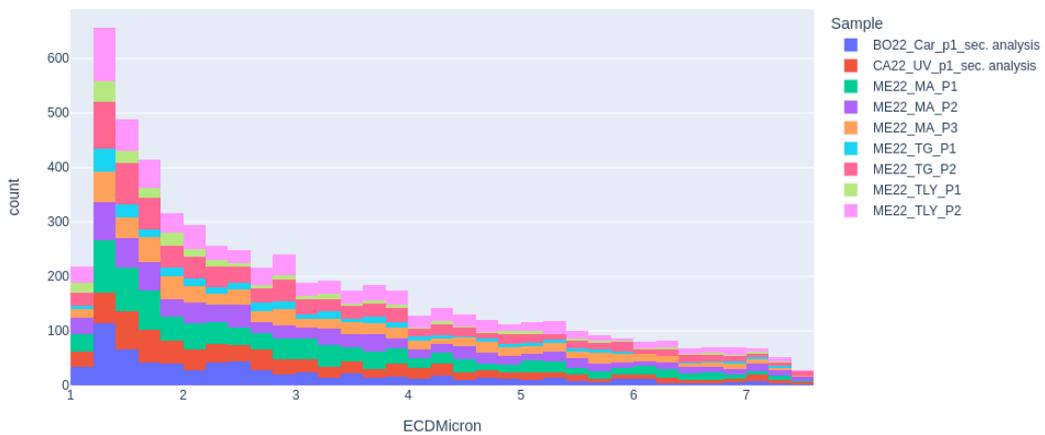


Figure 3.7: Distribution of particle size per sample.

Furthermore, the bar plot displaying the total volume of particles per group highlights an interesting trend. In almost all stations, composite tire wear particles account for the majority of the collected particle material mass. This indicates that the presence of tire wear particles thus not being predominant in number of the particles collected, surpasses other particle types in terms of both mass and volume. The prevalence of composite tire wear particles suggests their significant contribution to the overall particulate matter composition in the analyzed samples. This finding underscores the importance of considering composite tire wear particles as a substantial component of airborne particulate matter, emphasizing their potential impact on air quality and human health.

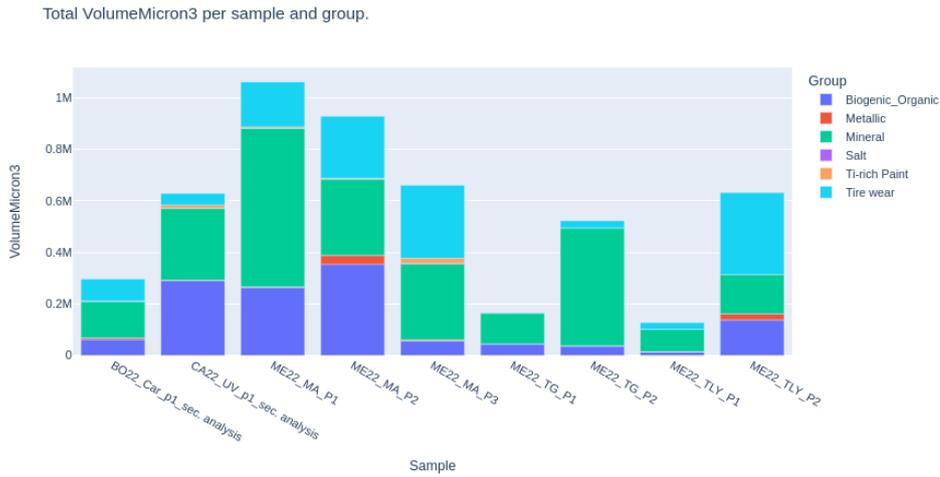


Figure 3.8: Contribution of total volume by group per sample.

Composition and Morphological Features: The analysis of the chemical composition of particulate matter revealed several key findings, the parallel bar graphs provide a comprehensive visualization of the composition of particles in the dataset, each vertical axis corresponds to the percentage of the element present in the sample, each particle line cross the vertical axis at the percentage of the element detected by EDS analysis, particles are grouped by color corresponding different groups to facilitate trend analysis. Organic particles were predominantly composed of carbon, with concentrations ranging from 50% to 90%, along with the presence of oxygen. Mineral particles, on the other hand, were primarily composed of (Si) silicon-oxygen crystals, although smaller quantities of calcium and aluminum were also detected. Metallic particles were consistently identified as iron (Fe), often with varying degrees of oxidation. Tire wear particles exhibited composite compositions with different ratios of carbon (C), oxygen (O), iron (Fe), aluminum (Al), silicon (Si), and calcium (Ca). Notably, salt particles were distinct, being the only type to contain sodium (Na) and chlorine (Cl). These findings provide valuable insights into the chemical composition and elemental constituents of different particle types within the analyzed particulate matter samples.

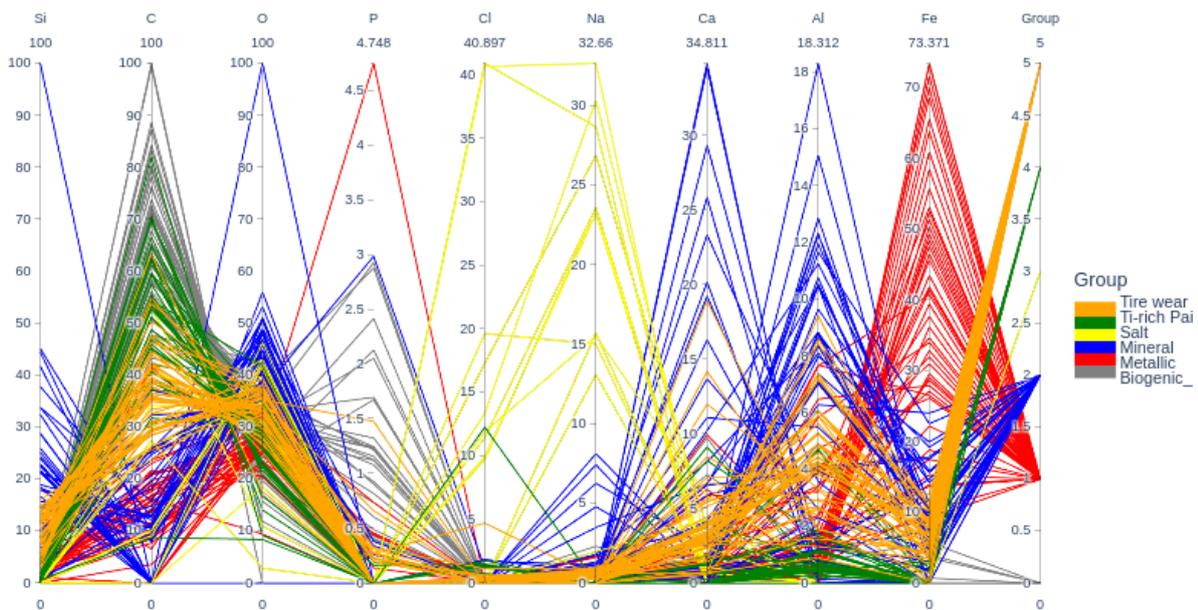


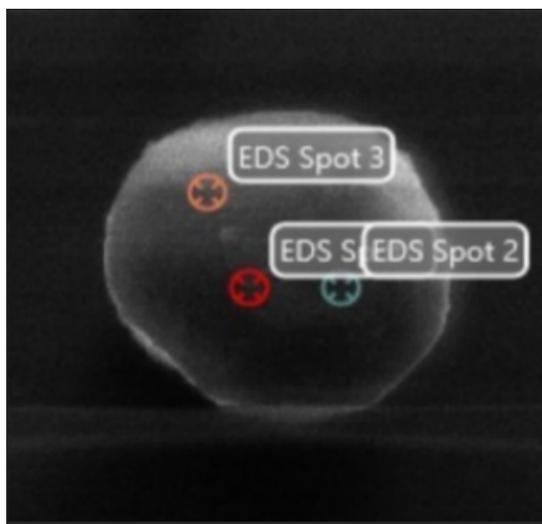
Figure 3.9: Elemental composition of particles per Group

Overall, the analysis of the database offers a comprehensive view of the particulate matter composition, including group distributions, size variations, and chemical composition. These findings provide valuable insights into the sources, characteristics, and potential impacts of the analyzed particulate matter, contributing to the field of environmental research and aiding in the development of strategies for pollution control and mitigation.

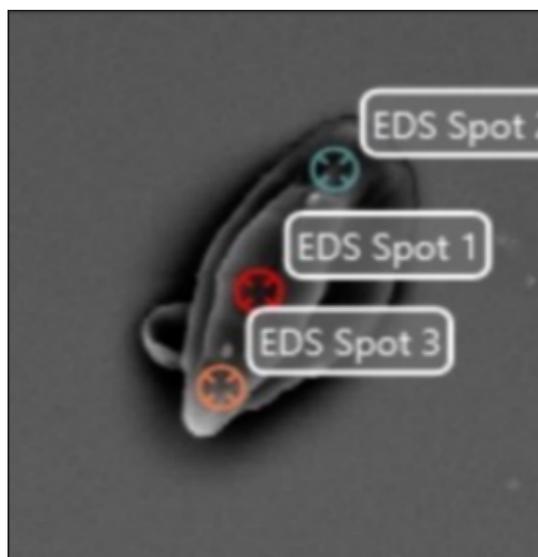
3.5. Origin Based Morphochemical groups.

3.5.1. Organic:

Organic particles are predominantly composed of carbon, with high concentrations from 50% to 90%, and oxygen. These particles are characterized by a diverse morphology, with some exhibiting wood like irregular shapes while others displays more regular structures like spores or seeds. The primary source of emission for organic particles is typically biogenic natural sources such as vegetation and organic matter decomposition. Organic particles are characterized by their high carbon content, which distinguishes them from other particle types. The presence of oxygen is also significant, often forming functional groups within the organic compounds. Organic particles can take various forms, including spherical, irregular, or chain-like structures depending on their chemical composition and formation mechanisms.



(a) Organic Particle Micrograph



(b) Organic Particle Micrograph

Figure 3.10: Micrographs examples for collected organic particles.

(a) Table 1

Element	Weight %	Atomic %	Error %
C K	60.43	68.93	5.68
O K	35.10	30.05	11.32
Si K	0.58	0.28	10.61
Zr L	0.59	0.09	18.49
Nb L	1.66	0.24	12.42
Ra M	0.60	0.04	35.79
K K	1.04	0.37	12.36

(b) Table 2

Element	Weight %	Atomic %	Error %
C K	76.27	82.97	5.47
O K	18.34	14.98	16.74
Na K	0.00	0.00	99.99
Mg K	0.64	0.34	12.27

Table 3.1: EDS analysis for both particles.

3.5.2. Minerals:

Mineral particles generally come from dust and secondary atmospheric chemical reaction products. If the particle displays irregular morphology, they usually come from fine particles in windblown dust on the Earth's surface. If they are regular, they come from chemical reactions in the atmosphere. Characteristic mineral particles usually show a mineral imprint (often silicates, calcite, or mixtures). They can be individual crystals or aggregates and common sources of these types of particles are road wear, construction work, quarrying and geogenic.

Consistent road wear particles appear in one or more bitumen-bound minerals (crushed grains), colorless or colored, in some cases pleochroic (eg biotite). Individual minerals (quartz, feldspar, pyroxene, amphibole, mica) and rock fragments (eg granite). Elements: Si, Al, Ca, Na, K, Mg; Fe, S. Origin: Abrasion of the road surface.

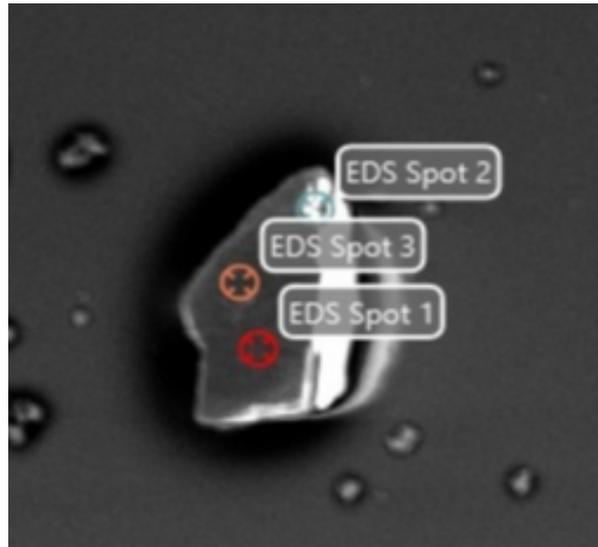


Figure 3.11: Mineral particle.

(a) Spot 1				(b) Spot 2			
Element	Weight %	Atomic %	Error %	Element	Weight %	Atomic %	Error %
C K	5.94	9.99	12.83	O K	56.29	70.55	7.17
O K	49.64	62.69	8.00	Na K	8.61	7.51	8.19
Na K	7.70	6.77	8.00	Br L	5.88	1.48	4.34
Br L	11.56	2.92	2.75	Al K	5.22	3.88	5.67
Al K	1.22	0.92	8.13	Si K	21.38	15.26	4.64
Si K	21.58	15.52	4.76	K K	0.44	0.23	15.64
Ca K	2.35	1.18	5.30	Ca K	2.18	1.09	5.54

Table 3.2: Combined Tables

3.5.3. Metals:

Particles containing mainly metals in its elemental composition (very often Fe ± Si, Cu, Cr, Sb, Sn, Ba, Zn, O, etc.). Metal particles produced by brake wear suspension are usually irregularly shaped metals with sharp edges and points. The most common source of these particles is the abrasion of brake pads, discs, and calipers. And with more often with the following elemental composition: Fe, Cu, Zn, Ti, Mo, Mn, Ba, Sn, W.



(a) Caption for Figure 1



(b) Caption for Figure 2

Figure 3.12: Caption for the entire figure

(a) Table 1

Element	Weight %	Atomic %	Error %
C	54.06	65.37	9.47
O	30.19	27.41	12.6
Al	2.47	1.33	8.51
Si	4.99	2.58	5.82
S	3.05	1.38	7.11
Fe	2.61	0.68	12.34

(b) Table 2

Element	Weight %	Atomic %	Error %
C K	54.06	69.53	10.32
O	20.2	19.5	14.68
Mg	1.65	1.05	14.46
Al	4.33	2.48	7.93
Si	7.91	4.35	5.99
Fe	3.71	1.03	14.6

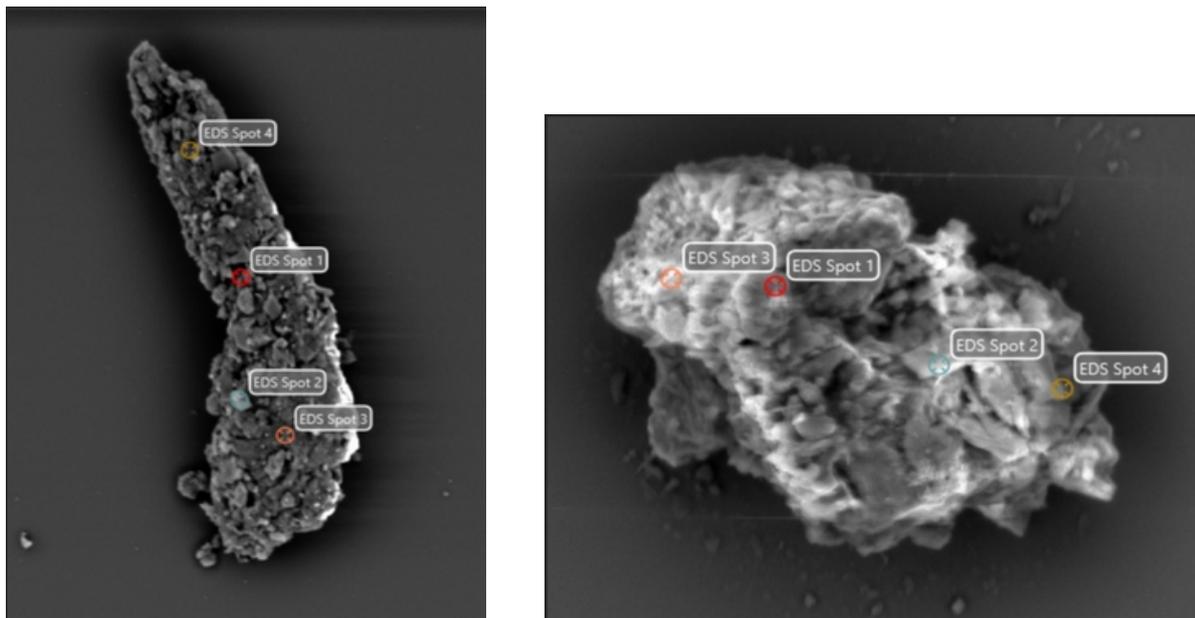
Table 3.3: Combined Tables

3.5.4. Composite:

Composite particles from tire wear typically originate from the friction and abrasion between vehicle tires and the road surface, as well as the brake discs. These particles exhibit a rounded or kidney-shaped morphology, with elongated and cylindrical structures resembling cigarette ashes. They range in length from 10 to 200 μm and have a width of less than 20 μm . The composition of tire wear particles is complex, consisting of a central core composed of worn rubber from the tire tread. This core can be partially or fully covered by smaller particles that originate from brake wear, road wear, and other road dust sources such as minerals and metallic particles.

Characteristics element traces are different if the particle is partially covered by road wear and the brakes: C, S \pm Zn; Si, Al, Na, Ca, K, Mg; Fe, Cu.

Or can be completely covered in road and brake wear: Si, Al, Ca, Na, K, Mg; Fe, Cu, Zn, Ti, Mo, Mn, Ba, Sn, W.



(a) Composite particle 1.

(b) Composite particle 2.

Figure 3.13: Micrographs examples of composite particles

(a) Micrograph of composite particle 1.

Element	Weight %	Atomic %	Error %
C K	26.64	39.29	9.46
O K	43.35	48.00	8.99
Na K	0.80	0.62	23.44
Mg K	1.85	1.35	8.68
Al K	7.22	4.74	5.22
Si K	6.15	3.88	5.28
W M	11.39	1.10	4.90
K K	0.50	0.23	21.12
Ca K	1.10	0.49	16.68
Fe K	1.00	0.32	22.90

(b) Micrograph of composite particle 2.

Element	Weight %	Atomic %	Error %
C K	0.37	0.70	99.99
O K	40.47	56.54	9.07
Na K	1.21	1.18	21.97
Mg K	2.32	2.13	10.51
Al K	10.41	8.62	5.67
Si K	27.62	21.98	4.67
K K	10.00	5.72	4.69
Ca K	1.52	0.85	21.44
Fe K	3.70	1.48	9.94
Zn K	2.38	0.81	14.34

Table 3.4: Micrographs of composite particles.

4

Standard Practice for particulate material Analysis by SEM-EDS.

4.1. Scope

- 4.1.1. This practice covers the analysis of airborne particulate material by scanning electron microscopy/energy-dispersive X-ray spectrometry (SEM/EDS). The analysis is performed by manual operation of both the SEM and EDS systems, to screen the sample for candidate particles that could be associated with any of the interest morphochemical groups.
- 4.1.2. Since software and hardware formats vary among commercial systems, guidelines will be offered in the most general terms possible. For proper terminology and operation, consult the SEM/EDS system manuals for each instrument.
- 4.1.3. The values stated in SI units are to be regarded as standard. No other units of measurement are included in this standard.

4.2. Terminology

- 4.2.1. Characteristic particles, *n* – particles that have compositions rarely found in particles from any sources other than the related morphochemical group of interest.
- 4.2.2. Consistent particles, *n* – particles that have compositions and morphological features that are found in the specific morphochemical group of origin but also arise from other non-identified sources.
- 4.2.3. Morphology, *n*; morphological, *adj* – referring to size, shape, structure, and texture.

4.3. Summary of Practice

- 4.3.1. Particles composed of high mean atomic number elements are detected by their SEM backscattered electron signals, and an EDS spectrum is obtained from each. The EDS spectrum is evaluated for constituent elements that could identify the particle as being consistent with or characteristic of any of the selected morphochemical groups (2-4). See Section 9 for discussion on classification of particles.

4.4. Significance and Use

- 4.4.1. This document will be useful for laboratory personnel involved in analyzing particulate material using SEM/EDS.
- 4.4.2. SEM/EDS analysis of particulate materials is a non-destructive technique that offers both morphological information and the identification of constituent elements in individual particles.

- 4.4.3. Particle analysis, unlike bulk sample methods such as atomic absorption spectrophotometry (AAS), neutron activation analysis (NAA), inductively coupled plasma atomic emission spectrometry (ICP-AES), and inductively coupled plasma mass spectrometry (ICP-MS), allows for the examination of the material without dissolving or extracting it. This non-destructive nature of particle analysis preserves the size, shape, and individual particle characteristics, enabling the detection and identification of different morphological groups within the particulate material.

4.5. Sample Preparation

- 4.5.1. Once the sample is removed from the box, care should be taken so that no object touches the surface of the collection substrate and the sample should not be left uncovered any longer than is reasonable for transfer, mounting, or labeling.
- 4.5.2. The sample substrate should be placed in a specimen stub adequate for high vacuum microscopy.
- 4.5.3. Since the substrate may contain plastics and other non-conductive materials, the sample will need to be coated to increase its electrical conductivity, unless an environmental SEM or variable pressure/low-vacuum SEM is used for the analysis. Gold alloys like gold-palladium are a good choice of coating material since it will not interfere with X-ray lines of interest in the case of organic particulates.
- 4.5.4. For high-vacuum SEM analysis, it is important to apply a sufficient coating to prevent sample charging. The coating process involves sputter coating, which involves depositing a thin layer of conductive material onto the sample surface. The two key parameters to consider during sputter coating are the coating time and the thickness of the coating layer (referred to as the coating thickness or coating thickness).
- 4.5.5. The following is a summary of the recommended procedure using the specified configuration:
- 4.5.5.1 Coating Material: Use a gold-palladium (Au-Pd) alloy as the coating material. It is chosen for its compatibility with SEM/EDS analysis and minimal interference with X-ray lines of interest, especially when analyzing organic particles.
 - 4.5.5.2 Sputtering Configuration: Set the sputtering parameters as follows:
 - 4.5.5.2.1 Current: Use a current of 20 mA to ensure proper deposition of the coating material on the sample.
 - 4.5.5.2.2 Vacuum: Maintain a vacuum pressure of 8 mbar throughout the coating process. Adequate vacuum conditions are crucial for achieving uniform and controlled deposition.
 - 4.5.5.2.3 Coating Time: Coat the sample for 120 seconds, allowing sufficient material deposition to achieve an optimal thickness ranging between 10-15 nm.
- 4.5.6. The sample collection stub shall be labeled in such a manner that it is distinguishable from other sample collection stubs without compromising the sample; for example, label the bottom or side of the stub.

4.6. Sample Area

- 4.6.1. Sample collection stubs for SEMs typically come in one of two diameters: 12.7 mm or 25.4 mm, which yield surface areas of 126.7 mm² and 506.7 mm² respectively.
- 4.6.2. Analysis of the maximum allowable surface area of the sample is recommended, however, many automated systems can be programmed to terminate the analysis of a stub or series of stubs once a pre-established number of particles have been detected. The decision as to how many particles satisfy the requirements of a particular case should be set out in the laboratory's standard operating procedures.

4.7. Instrument Requirements and Operation

4.7.1. General

- 4.7.1.1 Most commercial-grade SEM/EDS systems should be adequate for the particulate material analysis given that the criteria set forth in 4.7.2 and 4.7.3 are met.

4.7.2. Scanning Electron Microscope (SEM)

- 4.7.2.1 The SEM, operating in the secondary electrons imaging mode, shall be configured to visualize particles down to at least 1.0 μm in diameter. The recommended magnification is between X5000-X10000 to ensure a "panoramic view" of particles in the sample.
- 4.7.2.2 The SEM shall be capable of an accelerating voltage of at least 20 kV.

4.7.3. Energy Dispersive X-ray Spectrometry (EDS)

- 4.7.3.1 The detector shall be configured to produce a resolution of better (less) than 150 eV during analysis, measured or extrapolated as the full width at half the maximum height of the Mn Ka peak.
- 4.7.3.2 At a minimum, the EDS spectrum shall be acquired at 20 eV per channel.
- 4.7.3.3 Display of the EDS output shall encompass the X-ray lines of analytical utility, with a minimum range of 0–15 keV.
- 4.7.3.4 Automated systems will also include software capable of acquiring X-ray spectra for a specified collection time or total X-ray counts.
- 4.7.3.5 The instrument shall be capable of recording spectra obtained from the analysis of candidate particles at specific locations.
- 4.7.3.6 The instrument should be capable of capturing spectra with a specific number of counts.

4.7.4. Sample Placement

- 4.7.4.1 Verify that the sample has been correctly coated with a conductive material before placing it on the SEM.
- 4.7.4.2 Record the positions of the stubs on the SEM stage when the samples are inserted.
- 4.7.4.3 Though it is not necessary for the analysis of the sample and collections of the data, if additional analyses will be needed, it is desirable that the stub can be returned to the same orientation as before its removal. This could consist of marking the side of each stub and aligning it with marks on the microscope stage or by having stubs that fit into the stage in only one position (for example, stubs with a pin that is a half-circle in cross-section).

4.7.5. Sample Processing and Data Acquisition. Detection and Calibration

- 4.7.5.1 Particles that are visually differentiable at x500 magnification are candidates to be characteristic of any of the morphochemical groups of origin.
- 4.7.5.2 Particle imaging should be started from the left-topmost region of the substrate and sliced in a raster scan pattern until most of the substrate area is traveled.
- 4.7.5.3 Secondary electron micrographs should be taken when the image contains visually differentiable particles.
- 4.7.5.4 When micrographs are taken, EDS spectra could be captured for particles of special interest. Using the spot EDS function, place a spot over each particle of interest and capture the signal until X-ray count requirements are met. Be careful with the placement of the spot; if the EDS spot size is bigger than the particle, higher magnification is needed to ensure that the captured spectra do not contain chemical information from outside the particle (substrate).
- 4.7.5.5 SEM micrographs should be saved in PNG format with at least 1080x720pp definition.
- 4.7.5.6 EDS spectra should be saved in .CSV format.

4.8. Quality Control

When conducting analysis and collection of particulate material samples, special measures have to be chosen to meet common quality management demands. The use of control charts and/or tracking of instrument parameters and quality control metrics is recommended. Therefore, as minimum conditions:

- 4.8.1 Establish a protocol to confirm appropriate instrument operation parameters prior to sample analysis.
- 4.8.2 Monitor the EDS X-ray energy calibration and SEM beam current stability at least once per batch of samples using appropriate reference materials of known elemental composition such as copper, cobalt, etc.
- 4.8.3 Analyze a reference material with particles of known size range and composition to test the accuracy of particle detection and identification. It is recommended that the reference material has been prepared and mounted in a manner comparable to the collection method in use by the current experiment. The frequency of analysis of this sample shall be subject to guidelines set out in the laboratory's standard operating procedures.
- 4.8.4 The incorporation of environmental or control samples into the analytical protocol is recommended to monitor the cleanliness of the sample collection or analytical system, or both. An environmental sample can be prepared in several ways: for example, it may be an unused stub that has been prepared contemporaneously with the questioned samples or a sample taken from the sample collection or analytical environments (exposed to the air or as a direct sample collection from clean workspace), or both.
- 4.8.5 Care should be taken when transferring samples to and from the instrument to prevent cross contamination. Tools should be cleaned using a solvent and lint-free tissue between samples from different sources.

5

Particle detection.

5.1. Particle detection by binary threshold.

Particle detection in SEM images can be approached as a problem of object recognition, where the goal is to identify and localize particles of interest in the image. Particle detection is a challenging task as particles can be irregularly shaped and small in size, making it difficult to distinguish them from the background. Furthermore, particles can be deposited not only on smooth surfaces but also in different kinds of structures with roughness or inside nanofiber structures, further complicating the detection process. Therefore, to effectively detect particles deposited in different types of substrates, it is essential to apply advanced techniques that can accurately separate them from the background and account for their irregular shape and location. However, assuming the substrate surface to be regular and smooth, as in this case, binary thresholding is a powerful technique that can help solve this problem.

Binary thresholding involves dividing an image into two regions based on an intensity threshold value. Pixels with intensities above the threshold value are classified as foreground, while pixels with intensities below the threshold value are classified as background. In the context of particle detection, binary thresholding can be used to separate particles from the background by setting an appropriate threshold value based on the intensity of the particles regions. One of the advantages of binary thresholding is that it is a simple and computationally efficient technique that can be easily implemented in many image processing software packages. It is particularly useful in cases where the objects of interest have a distinct intensity or color compared to the background, which is often the case for particles on smooth surfaces while visualized on SEM images as show on figure 5.1. Additionally, binary thresholding can be used for other applications such as image segmentation, object detection, and image enhancement.

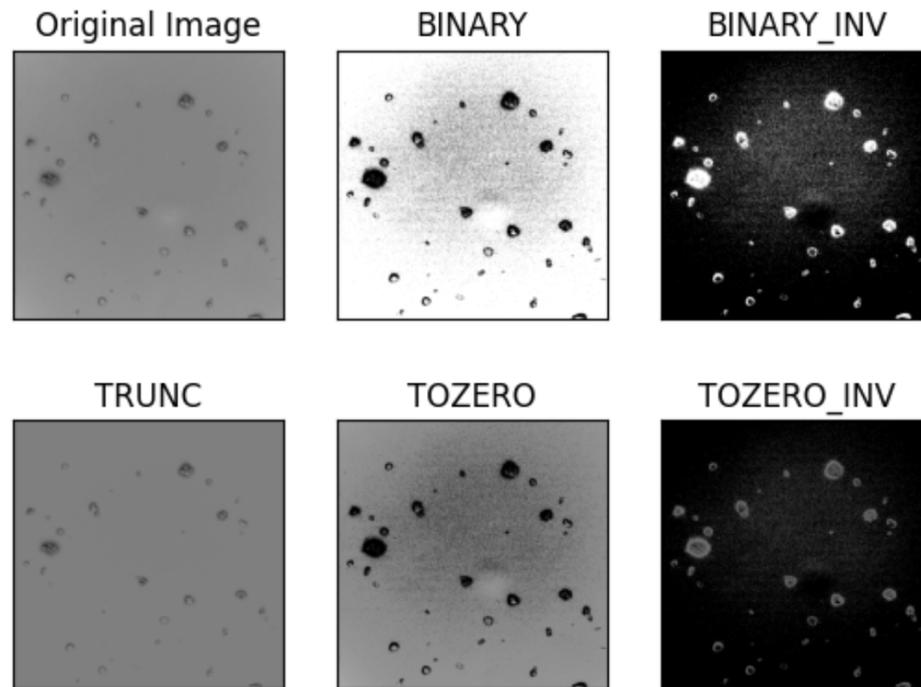


Figure 5.1: Original image and the results of thresholding using the following techniques: BINARY, BINARYINV, TRUNC, TOZERO, and TOZEROINV.

However, binary thresholding may not be suitable for object detection in cases where the objects of interest have a similar intensity or color to the background. In such cases, setting an appropriate threshold value to separate the foreground objects from the background may not be feasible, as the intensity or color of the objects may overlap with that of the background. For example, if we have an image of a black cat on a black background, binary thresholding may not be suitable for detecting the presence of the cat, as the intensity of the cat may be similar to that of the background.

In such cases, more sophisticated techniques such as edge detection, template matching, or machine learning-based approaches may be required for object detection. Edge detection can be used to detect the boundaries of objects based on changes in intensity or color. Template matching involves matching a pre-defined template to the image to identify objects that match the template. Machine learning-based approaches involve training a model to recognize objects based on a large dataset of labeled images. These techniques can be more effective in cases where the objects of interest have similar intensity or color to the background.

5.2. High frequency noise removal.

Smoothing an image involves applying a filter kernel to reduce noise and enhance its visual quality. The kernel is a small matrix of numbers that defines how the filter will operate on the image. The most common type of kernel is the Gaussian kernel, which applies a weighted average to the pixel values in the image. This operation reduces the high-frequency components in the image, resulting in a smoother image. This process can be used to enhance particle detection with binary thresholding by removing small artifacts that appear due to brightness changes during the image capture process. These artifacts can be caused by various factors, such as variations in illumination produced by the high energy electron ray generating noise in the image. This small artifacts can be mistakenly detected as particles, leading to false-positive detections.

To remove these artifacts, the image can be smoothed to reduce high-frequency components. This operation makes it easier to detect particles by applying binary thresholding, resulting in more accurate detections. This technique is commonly used in image processing applications, such as microscopy and computer vision, to improve the accuracy of object detection.

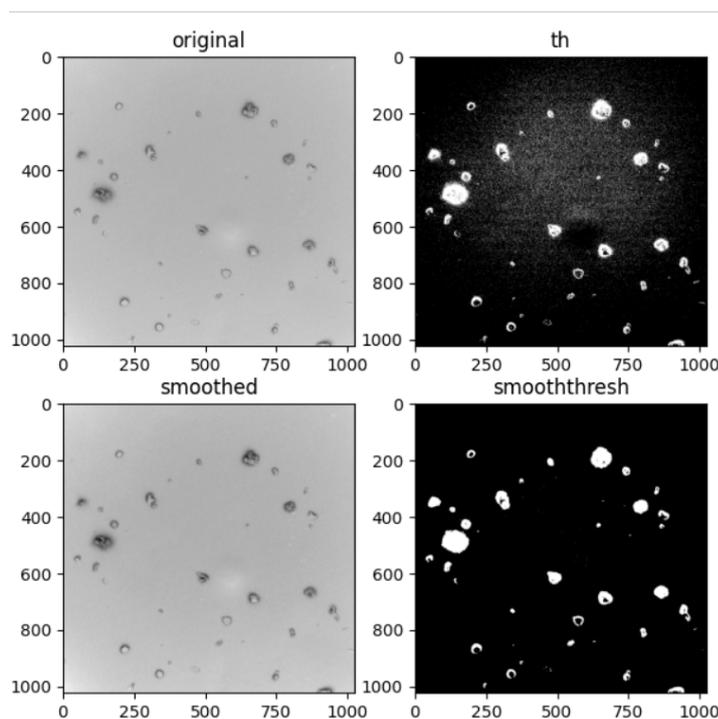


Figure 5.2: Image of original and smoothed SEM images with corresponding binary threshold applied. When smoothed, high frequency artifacts are removed.

5.3. Morphological operations on segmentation mask.

Morphological operations are fundamental techniques in image processing, they can be used to manipulate and extract features from an image. Two commonly used morphological operations are opening and dilation. Opening is an operation that involves the erosion of an image followed by dilation. It is commonly used to remove small artifacts or noise from an image while preserving the overall shape of larger structures. Dilation, on the other hand, involves expanding the boundaries of an object in an image.

In the context of particle detection, morphological operations can be applied to a segmentation mask to improve detection accuracy. A segmentation mask is a binary image that labels each pixel as either belonging to a particle or not. However, small artifacts or noise in the segmentation mask can lead to false-positive or false-negative detections, reducing the accuracy of the particle detection algorithm. In this case, dilation can be used to remove small artifacts, including extremely small particles that are not meant to be classified.

Dilation can also be used to fill holes and improve the segmentation mask. Holes in the segmentation mask can occur due to variations in the illumination or other factors that affect the binary threshold process. By applying dilation to the segmentation mask, the boundaries of the particles are expanded, and any holes or gaps between them are filled. This results in a more accurate and complete segmentation mask.

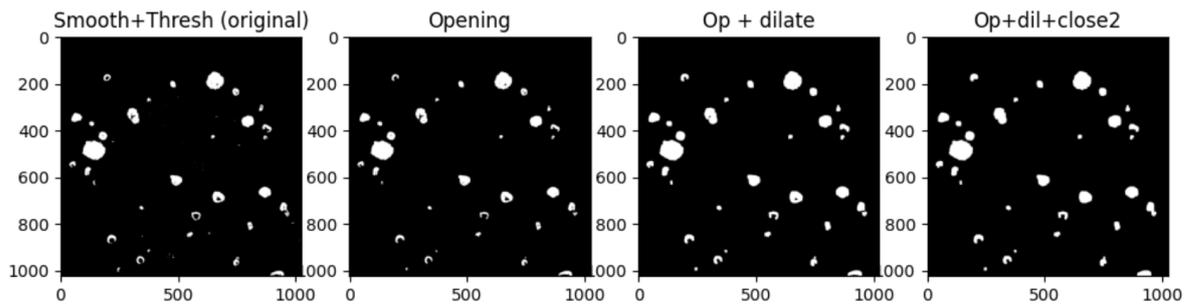


Figure 5.3: Morphology techniques to remove small artifacts and expand particles masks.

5.4. Connected region detection.

Regionprops is an image processing function that allows us to extract properties of connected components in binary images. In the context of particle detection, it can be used to extract properties such as area, perimeter, centroid, bounding box coordinates, and more from each particle detected in a segmentation mask.

The process behind regionprops involves first identifying all the connected regions in the binary image, which correspond to the individual particles in the case of particle detection. Then, for each connected region, a set of properties are calculated and returned in the form of a list or an array. These properties can be used to filter out small or large particles based on a specific range of areas or perimeters, or to calculate additional metrics such as particle size distribution or circularity.

One useful use case of regionprops for particle detection is to extract the bounding box coordinates of each particle, which can be visualized as rectangles overlaid on the original image. These bounding boxes can also be used to crop individual particle images from the original image, allowing for further single particle analysis as for the case of this work.

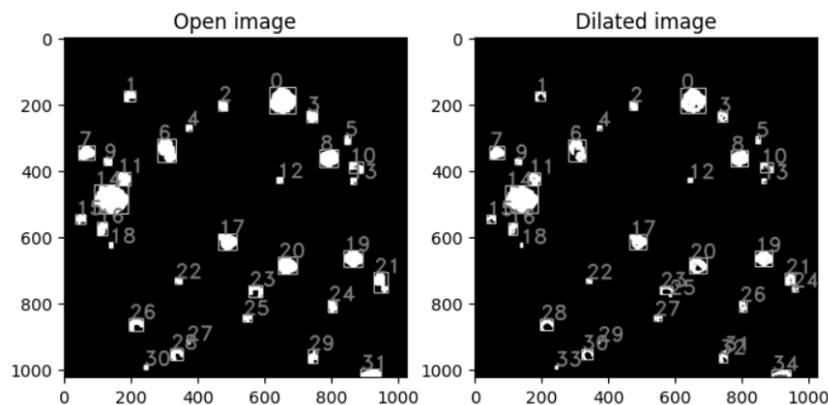


Figure 5.4: Detected bounding box for particles on image mask.

5.5. Single particle cropping.

Once the bounding boxes of particles in a segmentation mask have been detected, they can be used to crop individual images of each particle from the original image. This process involves obtaining the coordinates of the bounding box, which define a rectangular region that encloses the particle. These coordinates can then be used to extract the corresponding region of interest from the original image using image cropping techniques.

Cropping the individual images of particles is a crucial step in the particle detection process, as it allows for the isolation and analysis of each particle separately. By obtaining individual images, properties such as size, shape, and orientation can be measured and used for further analysis. Additionally, cropping the images can reduce the computational burden of processing the entire image, as only the relevant regions containing the particles need to be analyzed.

Once the individual images have been cropped, they can be further processed using various image analysis techniques to extract more information about the particles. This information can be used for a wide range of applications, including particle characterization, identification, and classification as in the next chapter.

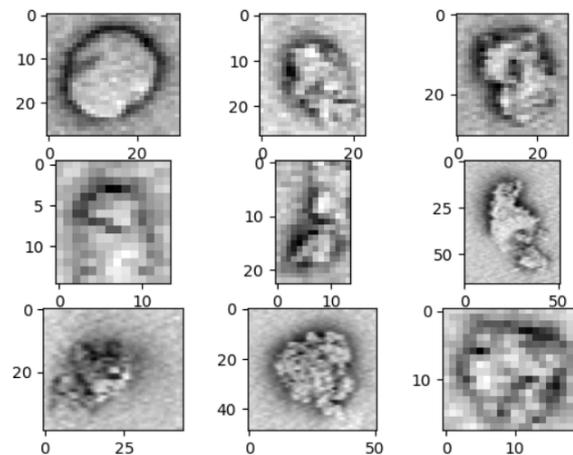


Figure 5.5: Cropped samples from original image and bounding box.

5.6. Code flow

The `detect_and_crop` function takes an input image and returns a list of cropped images. Here is the complete process performed by this function to detect and crop particles:

- 6.6.1 The `remove_sem_label` function is applied to the input image to remove the label in the bottom part, resulting in the SEM micrograph without the bottom label.
- 6.6.2 The `smooth_and_thresh` function is then used on the output image from the previous step. It applies a smoothing filter to reduce high frequency noise followed by a binary inverse threshold to obtain the raw segmentation mask for particles in the micrograph.
- 6.6.3 Next, the `apply_morp` function is applied to the output image from the previous step. This function performs morphological operations such as opening and dilatation to remove small artifacts and increase the quality of the segmentation mask.
- 6.6.4 The mask is further processed using the `detect_from_mask` function. This function detects connected objects in the binary image and returns a list of coordinates and bounding boxes.
- 6.6.5 Using the list of bounding boxes obtained from the previous step, the `crop_box` function crops the particles in the input image. It uses the bounding boxes from the binary mask detection and returns a list of cropped images.

6

Automatic Particle Detection.

6.1. Review of AI Techniques in SEM Image Segmentation

The application of artificial intelligence techniques in micrograph segmentation has shown promising results in recent research. AI techniques, such as deep learning and machine learning algorithms, have proven to be effective in automatically segmenting SEM images and highlighting specific features of interest. One popular approach in AI-based image segmentation is the use of convolutional neural networks. CNNs are capable of learning complex patterns and structures from large datasets, making them suitable for analyzing SEM micrographs. By training CNN models on annotated SEM images, researchers have been able to achieve high accuracy in segmenting various features like cracks, pores, and grain boundaries. Furthermore, the integration of AI techniques with traditional DIP methods has also shown promising results. By combining the strengths of both approaches, researchers have been able to enhance the accuracy and efficiency of SEM image segmentation. For example, incorporating pre-processing techniques like denoising and contrast enhancement using DIP algorithms can improve the quality of the input images.

The paper of [9] introduces a novel method utilizing computer vision and machine learning to analyze Scanning Electron Microscopy (SEM) images of nanomaterials. This approach proves effective even for complex shapes like overlapping nanoparticles, rod structures, and core-shell nanoforms. The results exhibit a strong correlation with manually measured sizes, with Spearman coefficients of 0.91 for the fully automated process and 0.99 for the semi-automated process. The main code implementation involved C++, QT, EAST, Tesseract, and OpenCV, and a CNN-based morphology classification algorithm was employed for evaluation. By comparing the proposed method with a baseline approach, the study demonstrated significantly improved performance in accuracy, precision, recall, and f1-score. The semi-automatic process showed enhanced accuracy compared to the fully automated one, and both methods exhibited high correlation between manual and algorithm-generated sizes.

In (Srikanth Sagar Bangaru, 2022) [10], the study leverages the U-Net architecture, originally designed for biomedical imaging, to analyze geological features in SEM images of shale samples. By employing a modified weight function, the model effectively discriminates between clay aggregates, matrix mineral particles, and organic matter. This addresses a significant challenge in image analysis where grey-level pixels are hard to distinguish. The research underscores the importance of understanding pore structure, rock properties, and fluid interactions in shale reservoirs for efficient resource extraction. Traditional segmentation methods struggle due to complex boundaries and low contrast, making deep learning, specifically U-Net, a promising tool for precise pixel-wise segmentation. The study provides comprehensive details about data selection, preprocessing, network training, and model validation. The results demonstrate improved segmentation accuracy compared to traditional methods, showcasing the feasibility and potential of texture-based deep learning in geological analysis. The availability of source code on GitHub further promotes transparency and reproducibility in this field of study.

The work of (Zhuoheng Chen, 2020) [11] introduces a U-Net convolutional neural network-based model for automated concrete microstructure analysis using SEM images. This model effectively identifies aggregates, anhydrous cement, hydrated cement, and pores in concrete specimens, achieving a

performance accuracy of 94.43% in validation with SEM images from seven concrete specimens. The method is automatic, accurate, cost-effective, and adaptable to various SEM images. It addresses the limitations of current semi-automated methods, providing a promising solution for precise and efficient concrete microstructure analysis. The U-Net model's architecture and training details are thoroughly explained, demonstrating its effectiveness in segmenting concrete components. Additionally, the study compares the U-Net model's performance with a Random Forest classifier, confirming its superiority in microstructure analysis. Overall, the research showcases the potential of deep learning models in civil infrastructure and microstructural analysis, paving the way for broader applications in different specimens and magnifications.

The Segment Anything (SA) (Kirshick, 2023) [12] project presents a groundbreaking advancement in image segmentation, comprising three integral components: task, model, and data. The task, termed Promptable Segmentation, aims to generate accurate segmentation masks based on diverse prompts, encompassing spatial and text information. The model, known as Segment Anything Model (SAM), is composed of three core elements: image encoder, prompt encoder, and mask decoder, collectively delivering efficient mapping of inputs to segmentation masks. The dataset, SA-1B, is unprecedented in scale, boasting over 1 billion meticulously annotated masks from 11 million privacy-protected images. SAM is rigorously evaluated, exhibiting remarkable proficiency in generating high-quality masks across various tasks. It notably excels in zero-shot transfer protocols through prompt engineering. Importantly, the project addresses fairness concerns, ensuring equitable performance across different demographic groups. SA-1B and SAM are made accessible to the research community under an open license, further fostering progress in image segmentation.

6.2. Implementation of an automatic particle segmentation pipeline.

The Work of (Rühle, 2021) [13] leverages a novel methodology for training Deep Segmentation Models for automated SEM image analysis, directly addressing the main concerns of collecting well-annotated image data. To overcome this, Generative Adversarial Networks (GANs) are employed to generate training samples in the form of SEM images and segmentation masks, based on real Micrographs without the need of manual annotation. The workflow results in particle segmentation, but spans multiple steps starting with a GAN-based particle shape generation, CycleGAN for image-to-image translation and MultiRes UNet architecture for image segmentation, demonstrating commendable performance. The research places emphasis on the quantitative descriptors that could be calculated from the segmentation mask, such as minimum Feret diameter, for nanoparticle classification and the outcomes are compared across different segmentation methods, with meticulous evaluation metrics like the Jaccard index. The whole code is an open-source implementation written in python using the TensorFlow DNN framework. The GAN, Cycle-Gan and UNet implementations are based on the keras models of the same name. GitHub repository is available for the code and special customization need to be done in order to extend to different datasets.

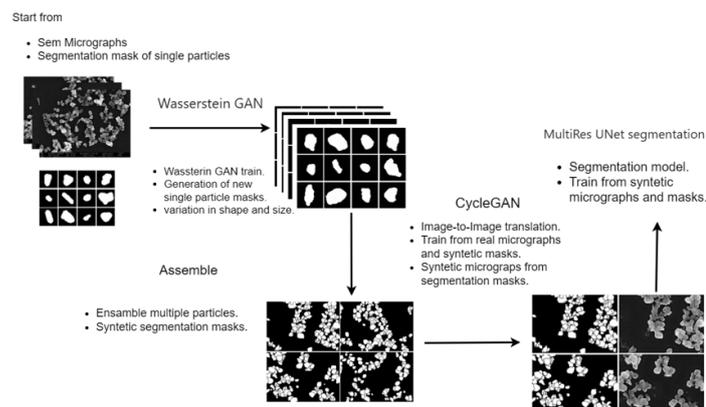


Figure 6.1: Process for automatic particle segmentation in SEM images.

The process starts from a set of Micrographs with particle samples captured with a SEM microscope and a collection of Single particles masks with variety of shapes and features. In the first step, a

Wassterian GAN model is trained for 1000 epochs over the set of masks with particle shapes, the training process is made with a batch size of 64 images. After the training process, the generator model is used to synthesize 1000 new files with a high diversity of particle shapes. These particles shapes are then arranged in a Perlin noise distribution with the noise threshold and noise frequency at 0.5 and 4 respectively, a minimum of 3 particles and maximum of 20 particles are also set in the configuration for generating fake segmentation mask with high-density clusters with different sizes similar from the mask of Micrographs manually annotated.



Figure. Particle shapes generated from Wassterian GAN.

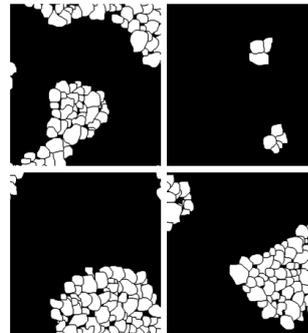


Figure. Agglomerated Synthetic masks.

Figure 6.2: Particle shapes generated from Wassterian GAN

In the next step, an implementation of Cycle-Gan is trained for image-to-image translation between Micrographs domain (A) and segmentation mask domain (B). The model is initialized with 64 filters and trained during 50 epochs on batches of two images.

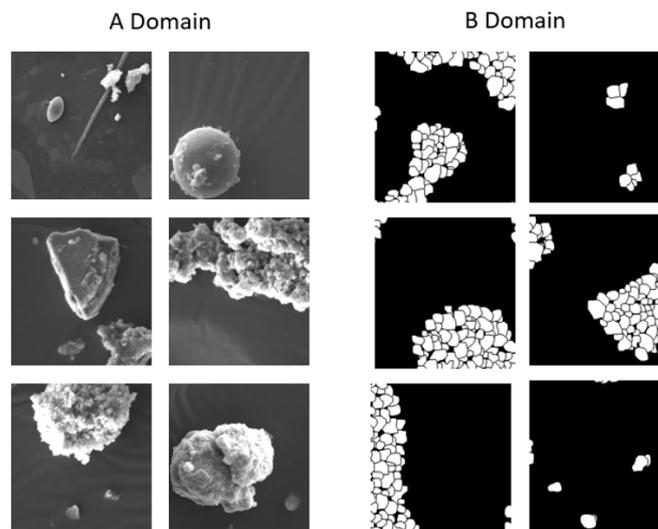


Figure 6.3: Image samples from Cycle-GAN training, synthetic segmentation mask in A Domain and real micrographs in B domain.

The training process led to the creation of a generator model capable of synthesizing fake SEM micrographs of particles from segmentation masks.

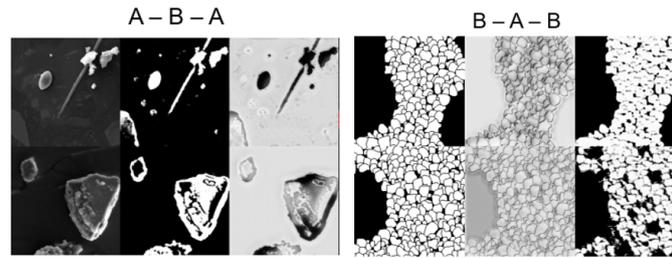


Figure 6.4: Image-to-Image translation from cycle-GAN. A to B translation generate mask from SEM micrographs. B to A translation generate synthetic micrographs from segmentation masks.

In the final steps, a large set of fake SEM micrographs are generated from the synthetic segmentation mask arranged in the previous steps, this pair of images collection is then used as a training set for a MultiResNet UNet implementation. The training process is made during 50 epochs with a batch size of 2 images.

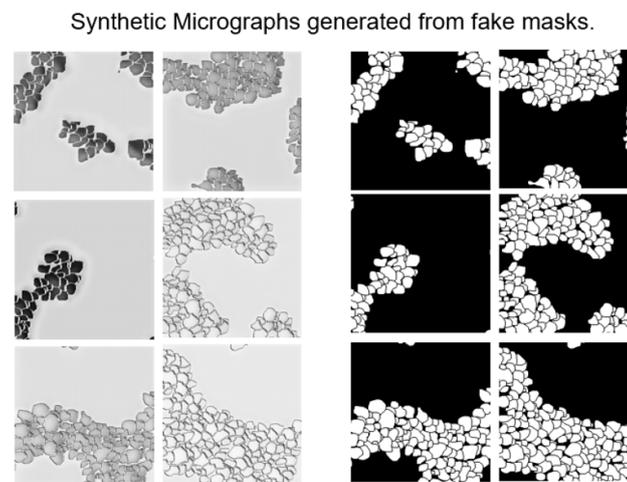


Figure 6.5: Training images for UNet implementation. In the left, synthetic particle micrographs. To the right, aggregated segmentation masks.

Inference is then performed over the real Micrographs, the cycle GAN translator generates segmentation mask with aggregated shapes, in the case of UNet segmentation model, the results display smaller and sparser shapes.

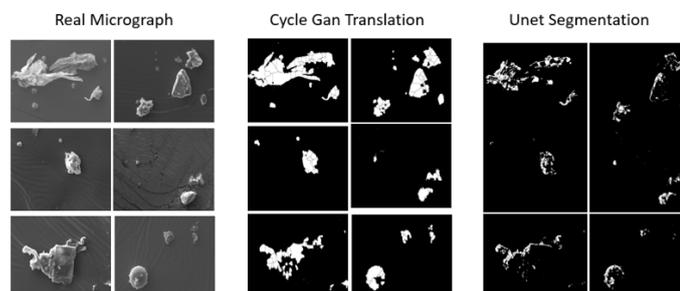


Figure 6.6: Results for the automatic segmentation training process.

6.3. Conclusions on AI and SEM Image Segmentation

Cycle-Gan translated segmentation masks display remarkable results that can be improved using DIP techniques like morphology operations and filtering techniques, this post processed mask can be used

as final segmentation mask for particle detection. The entire process was executed utilizing 50 authentic SEM micrographs and ran seamlessly on an NVIDIA RTX 2080 Ti, completing within approximately 48 hours. The results are promising, indicating the potential to train effective particle segmentation models without the need for labor-intensive manual annotations. This approach proves valuable, especially when confronted with the challenge of training a singular model for segmentation images across different microscopes. The varying diapositives and configurations inherent to different devices often result in substantial differences in brightness, contrast, and overall imagery. Through this process, specialized models can be crafted to cater to specific needs, such as accommodating different devices and substrates. This targeted approach ensures that the models are finely tuned to the unique characteristics of diverse microscopy and sampling setups, offering a pragmatic solution to the complexities posed by variations in imaging conditions.

7

Vision particle classifier.

7.1. Training vision model

In the deep learning experiment, a vision model was trained to classify particles into different groups, namely BiogenicOrganic, Mineral, Metal, and TireWear Composites. The experiment utilized three different models: Resnet50 with random weights, Resnet50 with ImageNet weights, and miniXception.

First, the dataset consisting of labeled particle images was prepared for training, validation, and testing. The images are preprocessed on runtime by resizing them to a consistent resolution for the model and normalizing pixel values. Data augmentation techniques such as random rotations, flips, zooms and random equalization were also applied to enhance the model's ability to generalize.

The models were initialized with the specified weights. The resnet50 model with random weights started with randomly initialized parameters, while the resnet50 model with ImageNet weights utilized pre-trained weights from the ImageNet dataset. The miniXception model, a compact and efficient architecture for low resolution images, was also used for comparison. During the training phase, the models were trained on the labeled particle images using a suitable loss function such as categorical cross-entropy. The training process involved iterative optimization of model parameters by adjusting the weights based on the gradients computed during backpropagation. The training was performed on a large number of epochs to allow the models to learn the intricate patterns and features associated with each particle group.

During the training process, validation was crucial to monitor the models' progress and prevent overfitting. A separate validation set, distinct from the training set, was used to evaluate the models' performance at regular intervals. This validation set contained labeled particle images that were not seen by the models during training. The models' predictions on the validation set were compared to the ground truth labels. Evaluation metrics such as accuracy were calculated to assess the models' reliability. These metrics provided valuable insights into the models' performance and helped in determining whether any adjustments were needed, such as tuning hyperparameters or applying regularization techniques, to improve the models' accuracy and generalization. Furthermore, the validation results guided the training process by allowing for early stopping and lr reduction if the models' performance on the validation set reached a plateau or started to deteriorate. This helped prevent overfitting, where the models become too specialized in the training data and fail to generalize well to unseen data.

The experiments were repeated for each model variant, including resnet50 with random weights, resnet50 with ImageNet weights, and miniXception. This allowed for a comprehensive comparison of their classification performance and the determination of the most effective model for the task.

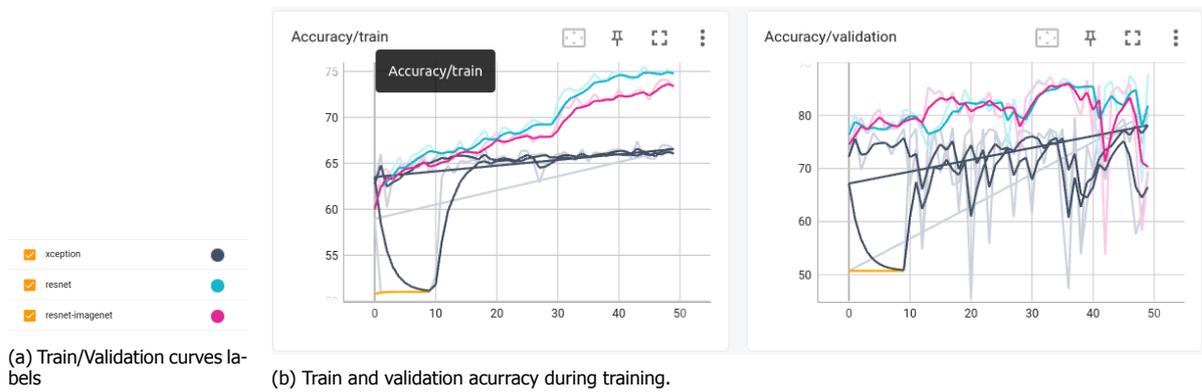


Figure 7.1: Micrographs examples for collected organic particles.

7.2. Model metrics and validation.

After training, the random weights initialized resnet50 shows the best performance on the validation set. The model were evaluated on a separate test set to assess their performance in classifying particles into the correct groups. Evaluation metrics such as accuracy, precision, recall, and F1 score were calculated to measure the models' performance.

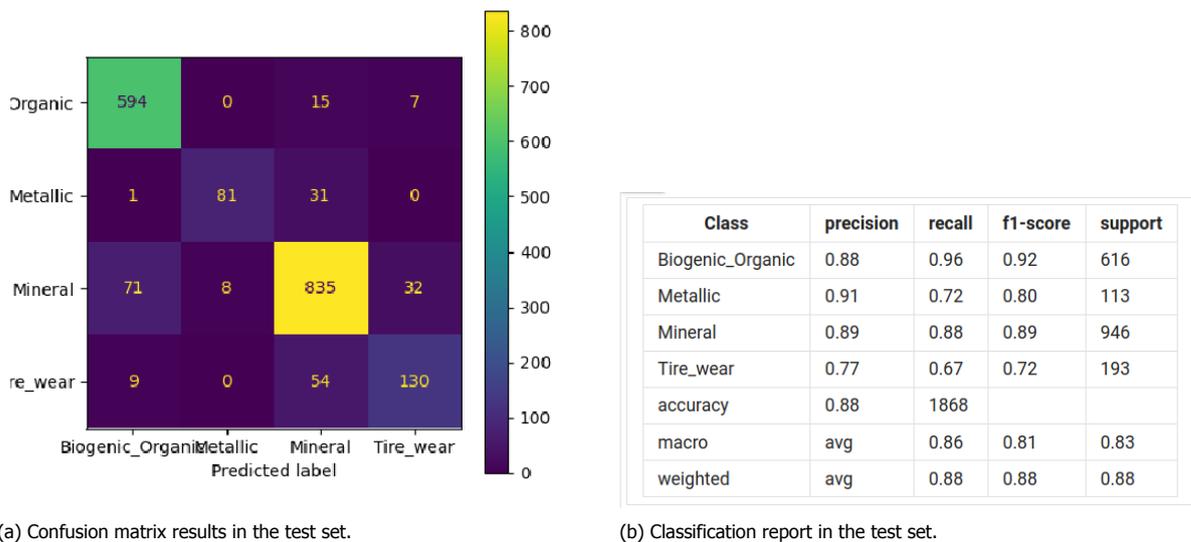


Figure 7.2: Classification results for RandomWeights Resnet50 model in the test set.

The accuracy metric, with a value of 0.88, indicates the overall proportion of correctly classified particles across all classes. Precision measures the ability of the model to correctly identify positive instances for each class. Recall, also known as sensitivity, represents the ability of the model to capture all positive instances. The F1-score combines precision and recall into a single metric, providing a balanced evaluation of the model's performance.

Analyzing the class-specific metrics, the BiogenicOrganic class achieved a precision of 0.88, indicating that 88% of the particles predicted as BiogenicOrganic were indeed correct. The recall of 0.96 suggests that the model successfully captured 96% of the actual BiogenicOrganic particles present in the dataset. The F1-score, which considers both precision and recall, was 0.92, reflecting a high overall performance for this class.

The Metallic class exhibited a higher precision of 0.91, indicating that the model was effective in identifying Metallic particles. However, the recall of 0.72 suggests that the model missed some Metallic particles in the dataset. Consequently, the F1-score for this class was 0.80, indicating a moderate level of performance.

The Mineral class showed a precision of 0.89, indicating a high proportion of correctly identified Mineral particles. The recall of 0.88 suggests that the model successfully captured 88% of the actual Mineral particles. The F1-score for this class was 0.89, indicating a balanced performance between precision and recall.

For the Tirewear class, the precision was 0.77, indicating that the model correctly identified 77% of the Tirewear particles. The recall of 0.67 suggests that the model missed some Tirewear particles in the dataset. The F1-score for this class was 0.72, indicating a moderate overall performance.

The macro-averaged metrics provide an average performance across all classes, giving equal weight to each class. The macro-averaged precision, recall, and F1-score were 0.86, 0.81, and 0.83, respectively. The weighted-averaged metrics take into account class imbalance by considering the support (number of instances) for each class.

7.3. Model analysis and conclusions.

Overall, the experiment yielded promising results with an accuracy of 0.88, indicating that the vision model was able to classify particles with a high degree of accuracy. The model demonstrated particularly strong performance in classifying BiogenicOrganic and Mineral particles, achieving high precision, recall, and F1-scores. However, there is room for improvement in accurately identifying Metallic and Tirewear particles, as indicated by slightly lower recall and F1-scores for these classes.

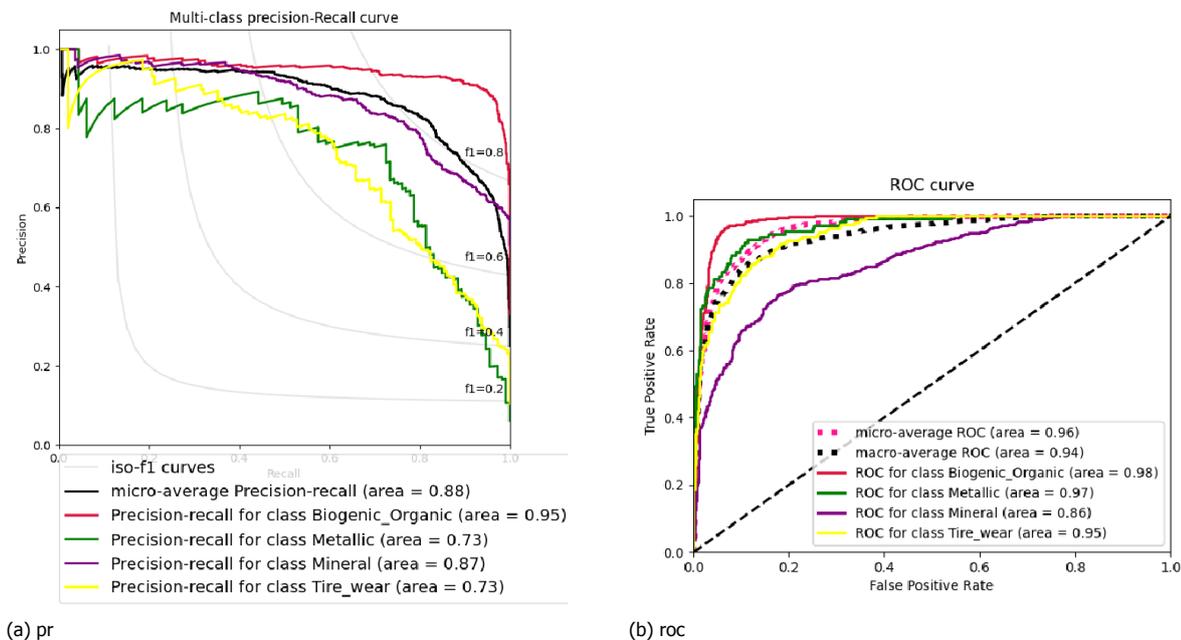


Figure 7.3: xxx

The Precision-Recall (PR) curve provides insights into the trade-off between precision and recall for different classification thresholds. In our experiment, the PR curve displayed a micro average Area Under the Curve (AUC) of 0.88 for all classes. This indicates a strong overall performance in capturing both positive instances and minimizing false positives across all classes. Examining the individual class results, the Organic class achieved a high PR AUC of 0.95, indicating excellent precision and recall for identifying Organic particles. The Metallic class had a PR AUC of 0.73, suggesting a moderate performance in distinguishing Metallic particles. The Mineral class exhibited a PR AUC of 0.87, indicating a good balance between precision and recall. The Tirewear class had a PR AUC of 0.73, suggesting a moderate performance in identifying Tirewear particles.

On the other hand, the Receiver Operating Characteristic (ROC) curve assesses the trade-off between the true positive rate (sensitivity) and the false positive rate (1 - specificity). The ROC curve for our experiment showed strong results, with an AUC of 0.98 for Organic particles, indicating high sensitivity and low false positive rate. The Metallic particles achieved a ROC AUC of 0.97, demonstrating

good discriminative power. The Mineral class had a ROC AUC of 0.86, indicating a reasonable performance in distinguishing Mineral particles. The Tirewear class displayed a ROC AUC of 0.95, suggesting good sensitivity and specificity. The micro-average AUC, calculated by aggregating the results across all classes, was 0.98 for the ROC curve, indicating an excellent overall performance in classifying particles. The macro-average AUC, which provides an average performance across classes, was 0.94, reflecting a strong discriminative ability.

In summary, the PR curve and ROC curve results indicate that the vision model performed well in classifying particles across different groups. The high micro-average AUC for the PR curve and ROC curve demonstrates the model's ability to effectively differentiate between particle classes. The specific AUC values for each class further highlight the varying performance levels, with Organic particles showing the highest discriminative power. These results validate the effectiveness of the trained model and provide valuable insights for particle classification tasks in various domains.

In conclusion, the deep learning experiment aimed to train a vision model for particle classification using different models, including resnet50 with random weights, resnet50 with ImageNet weights, and miniXception. The performance of these models was evaluated based on their ability to accurately classify particles into the BiogenicOrganic, Mineral, Metal, and TireWear Composites groups. The results of indicated that Resnet50 model were the most reliable between the model architectures tested during the experiment demonstrating a high classification capacity within the provided classes.

8

Semiparticle Analyzer: A Web Application for SEM Micrograph Analysis

Semiparticle Analyzer is a versatile web application designed for the in-depth analysis of scanning electron microscopy (SEM) micrographs containing particle samples. This tool is invaluable for researchers and professionals working in fields where the characterization of particle morphology is crucial.

Key Features:

Upload and Analyze: Users can effortlessly upload SEM micrographs through a user-friendly interface. The application processes the images, extracting individual particles for subsequent analysis.

Morphological Feature Calculation: The heart of Semiparticle Analyzer lies in its ability to compute essential morphological features for each particle. Parameters such as area, diameter, and other shape descriptors are calculated with precision, providing quantitative insights into particle characteristics.

Particle Classification: Leveraging machine learning techniques, the application goes beyond basic analysis. It employs a trained model to classify particles into distinct groups—such as Organic, Metal, Mineral, and Agglomerate—enabling a more comprehensive understanding of the sample composition.

Python-Powered Backend: Semiparticle Analyzer is built with Python, utilizing powerful libraries such as OpenCV for image processing and Torch for efficient image classification. This ensures robust performance and accurate results.

Streamlit for User-Friendly Demo: The front-end of the application is crafted with Streamlit, a Python library for creating web applications with minimal effort. This choice enhances user experience, making it intuitive for both experts and those less familiar with SEM analysis.

Binary Threshold Detection: The application employs advanced techniques, including binary threshold detection, to enhance the accuracy of particle identification within SEM micrographs.

Applications:

Semiparticle Analyzer finds utility across various scientific domains, including materials science, environmental science, and nanotechnology. Researchers can swiftly analyze large datasets, accelerating the pace of discovery and understanding in their respective fields.

Future Developments:

As a dynamic project, Semiparticle Analyzer will continue to evolve. Future enhancements may include expanded classification models, additional morphological features, and integration with cloud-based storage for seamless collaboration.

In conclusion, Semiparticle Analyzer stands as a powerful and accessible tool for scientists and researchers seeking a robust solution for SEM micrograph analysis. Its combination of user-friendly design, accurate feature calculation, and machine learning-driven classification positions it as a valuable asset in the realm of particle characterization.

Bibliography

- [1] T. Fisher, "Scanning electron microscopes," 2024. Accessed on 23 Jan. 2024.
- [2] S. Marks, "Introduction to energy dispersive spectroscopy (eds)," May 19 2020.
- [3] I. Colbeck and M. Lazaridis, "Aerosols and environmental pollution," *Naturwissenschaften*, vol. 97, pp. 117–131, Feb. 2010.
- [4] K. Shandilya, A. Gupta, and M. Khare, "Defining aerosols by physical and chemical characteristics," *Indian Journal of Air Pollution Control*, vol. IX, pp. 107–126, 01 2009.
- [5] N. Künzli, R. Kaiser, S. Medina, M. Studnicka, O. Chanel, P. Filliger, M. Herry, F. J. Horak, V. Puybonnieux-Textier, P. Quénel, J. Schneider, R. Seethaler, J.-C. Vergnaud, and H. Sommer, "Public-health impact of outdoor and traffic-related air pollution: A european assessment," *The Lancet*, vol. 356, pp. 795–801, Sept. 2000.
- [6] A. S. Pipal, A. Kulshrestha, and A. Taneja, "Characterization and morphological analysis of airborne pm2.5 and pm10 in agra located in north central india," *Atmospheric Environment*, vol. 45, no. 21, pp. 3621–3630, 2011.
- [7] J. Rausch, D. Jaramillo-Vogel, S. Perseguers, N. Schnidrig, B. Grobéty, and P. Yajan, "Automated identification and quantification of tire wear particles (twp) in airborne dust: Sem/edx single particle analysis coupled to a machine learning classifier," *Science of The Total Environment*, vol. 803, p. 149832, 2022.
- [8] J. D. Avellaneda, "Diferenciación del material particulado pm10-2.5 en el valle de aburrá (medellín, colombia) y asignación a sus fuentes: Aprendizaje del análisis sem/edx de partículas individuales,"
- [9] H. Kim and A. Initials], "Machine vision-driven automatic recognition of particle size and morphology in sem images," *Nanoscale*, 2020.
- [10] S. S. Bangaru and F. i. C. W. [Last name, "Scanning electron microscopy (sem) image segmentation for microstructure analysis of concrete using u-net convolutional neural network," *Automation in Construction*, 2022.
- [11] Z. Chen and F. i. X. L. [Last name, "Deep learning-based method for sem image segmentation in mineral characterization, an example from duvernay shale samples in western canada sedimentary basin," *Computers & Geosciences*, 2020.
- [12] A. K.-Y. Girshick, "Segment anything," 2023.
- [13] B. K. Rühle, "Workflow towards automated segmentation of agglomerated, non-spherical particles from electron microscopy images using artificial neural networks," *Scientific Reports*, 2021.