NFSU JOURNAL OF LAW & ARTIFICIAL INTELLIGENCE

JLAI
Volume: 1, Issue: 1
July 2022
E-ISSN: 2584-1661

MICROPLASTIC DETECTION IN THE ENVIRONMENTAL MATRIX USING ARTIFICIAL INTELLIGENCE: REVIEW OF RECENT ADVANCEMENT

Adjama Irédon*, Dr. Hemen Dave**

Abstract: In recent times, the presence of microplastic emerged as a serious environmental threat rendering ecological risk and human health hazard. Contemporary research indicates that microplastic is omnipresent in the environment including terrestrial, aquatic, aerial, and even biological environments i.e., within organisms as well as the human body. Therefore, detection and analysis of microplastics in the environmental matrix is a decisive task, which is further a necessity for the prevention and removal of microplastic pollution. However, microplastics arise from diverse sources and are of diverse types, and need to be detected for a wide variety of environmental matrices. Thus, to understand the level of microplastic pollution in the environmental matrix, the complexity associated with microplastic detection and analysis which include qualitative and quantitative detection followed by classification of microplastic according to the type of polymer, size, and shape, structure types (fiber, fragment, film), etc. Microplastic pollution in the environmental matrix is assessed either by microscopy and visual sorting or by spectroscopy. Many researchers have developed methods of visual detection using microscopes which are generally easy to apply but require a lot of human work time and are likely to reveal misleading results with a lack of further information on types of microplastics. While spectroscopy is a simple method to apply to a large number of samples, further complexity is associated with classifying microplastic. To solve these problems, scientists have resorted to the application of artificial intelligence (AI) for better detection and classification of the different types of microplastics accumulated in samples taken from various ecosystems during the last decades. Integration of AI with microscopic or spectroscopic detection of microplastic can be a forensic tool for microplastic detection to reduce the complexity associated with detection and identification. Machine learning or Artificial Neural networks can be a powerful tool for processing the images obtained by spectroscopy or microscopy for automatic and fast screening/classification of microplastics. AI-based detection of microplastic pollution in environmental matrix opens a new scope for big data processing with interpretability to provide reliable results and prediction. This study review methods for detecting microplastics in the environmental matrix using AI developed by researchers to automate and accurate categorization of microplastics in the environment.

Keywords: Microplastics, Detection, Environmental Matrix, Artificial Intelligence, Machine Learning, Forensic Tool

I. INTRODUCTION

Plastic materials are made of multiple varieties of polymers chosen along with color and other additive chemicals to provide desired properties as per applications and user requirements. The Organization for Economic Co-operation and Development (OECD) reports that global plastic production has increased from 234 million tonnes in the 2000s to 460 million tonnes in 2019s[1]. Today, the amount of plastic produced in the world today is 170 times higher than it was 60 years ago. At the same time, the production of plastic waste more than doubled, to 353 Mt in 2019. In 2019 alone, 22 million tons of plastic were released into the environment, including[2] water, lakes, and

* *School of Doctoral Studies and Research*, National Forensic Sciences University, Gandhinagar, Gujarat. The author can be reached at hemen.dave@nfsu.ac.in

** *School of Pharmacy*, National Forensic Sciences University, Gandhinagar, Gujrat. The author can be reached at hemen.dave@nfsu.ac.in

[1] Lenzen, M., et al, *The environmental footprint of health care: a global assessment*, The Lancet Planetary Health 271-279 (2020).

[2] Gambarini, V., et al, *Phylogenetic Distribution of Plastic-Degrading Microorganisms*. M-Systems, 12-20 (2021). Also see Kumar, R., & Sharma, P. *Recent Developments in Extraction, Identification, and Quantification of Microplastics from Agricultural Soil and*

oceans. Plastics represent "at least 85% of total marine litter", says the United Nations Environment Assembly. Due to low or slow degradation in environmental media, plastic waste can remain more or less inert for century years[3]. Thus, these accumulated in natural ecosystems and became emerging pollutants [4]. Under the effect of different variables, climates, and human activities, plastic waste is cut into small pieces of varying sizes. Plastic particles are generally, according to their size, classified into nano plastics (NPs) (those whose size is less than or equal to 100 nm) and microplastics (MPs) (those whose size is between 5 mm and 100 nm) [5].

NPS/MPs are less numerous and easily transported by water and wind, so they are present in all ecosystems on planet earth and their small size allows them to be absorbed by the smallest organisms and therefore disperse and accumulate in the food chain of all living organisms on the planet and cause health problems through poisoning[6]. MPs/NPS pollution in the environment emerging as a significant hazard for health and the ecosystem. Given this danger of intoxication that MPs/NPs alert to organisms and living beings, around 8407 research works have been published in the journals of Scopus and Google scholar highlighting the presence of MPs/NPs in various nature ecosystems (Figure 1) of which about 78% are articles (Figure 2). But it was from the year 2014, that researchers and research institutes began to develop methods for detecting tiny plastic particles to be able to identify invisible plastic pollution (Figure 3) [7]. To this end, this field is relatively recent whose approaches for sampling, analysis, identification, and classification of MPs and NPs are under development because of the diversity in the chemical composition of plastic particles present in the environment, their varieties, sizes, mixing and interaction with other pollutants present in the environment and with the matrices of the environment (water, soil, sediments, biota or air). For this, advanced and significant feats of harmonization of sampling procedures and laboratory analysis have emerged in scientific documents. But the identification and classification of polymers remains more or less complex and requires the use of measurement and analysis methods according to the particle size fractions of the plastic particles [8].

Generally, the detection of MPs is done by the visual method using microscopes and Nile red staining is generally easy to apply, but requires a lot of human labor and is likely to reveal results with 70% error and a lack of information on the different types of particles. of plastic present in the sample and they are not able to identify the polymer. The sizes of the MPs likely to be detected by this method are greater than 100 µm [9].

For the detection of MPs and NPs, the literature has mentioned various detection and

Groundwater, Fate and Transport of Subsurface Pollutants 125–143 (2021). Pasquier, G., et al, *The Golden Method for Sampling Surface Water Microplastics in Aquatic Environments*. Frontiers in Environmental Science, 10. (2022). Song, J. H., et al, *Biodegradable and compostable alternatives to conventional plastics*. Biological Sciences, 2127–2139 (2009).

[3] Polymer Profiles: *A GuIde to the World's Most WIdely Used Plastics*. Advancing Materials (2016).
Ritchie, H., & Roser, M. *Plastic Pollution*. Our World in Data. (2018).

[4] Acharya, S., et al, *Microfibers from synthetic textiles as a major source of microplastics in the environment: A review.* Textile Research Journal, 2136–2156(2021). Also see Agathokleous, et al, *Ecological risks in a 'plastic' world: A threat to biological diversity?* Journal of Hazardous Materials, 417, (2021). Ainali, N. M., et al, *Microplastics in the environment: Sampling, pretreatment, analysis, and occurrence based on current and newly-exploited chromatographic approaches.* Science of The Total Environment, 794, (2021). Allouzi, M. M. An et al, *Micro (nano) plastic pollution: The ecological influence on the soil-plant system and human health.* Science of The Total Environment, 788, 147815. (2021).

[5] Azeem, I.,et al, *Uptake and Accumulation of Nano/Microplastics in Plants: A Critical Review.* Nanomaterials, 11, (2021)

[6]Ali, I., et al, Micro- *and nano plastics in the environment: Occurrence, detection, characterization, and toxicity – A critical review.* Journal of Cleaner Production, 313, (2021). Also see Azizi, N., et al, *The quantity and quality assessment of microplastics in the freshwater fishes: A systematic review and meta-analysis.* Regional Studies in Marine Science, 47, (2021). Beaurepaire, et al, *Microplastics in the atmospheric compartment: A comprehensive review on methods, results on their occurrence, and determining factors.* Current Opinion in Food Science, 159–168, (2021).

[7] Anger, P. M., et al, *Raman micro spectroscopy as a tool for microplastic particle analysis.* Trends in Analytical Chemistry, 214–226, (2018).
Dong, M., et al, *Automated analysis of microplastics based on vibrational spectroscopy: Are we measuring the same metrics?* Analytical and Bioanalytical Chemistry. 022-039, (2022).

[8] Anger, P. M., et al, *Raman microspectroscopy as a tool for microplastic particle analysis.* Trends in Analytical Chemistry, 214–226, (2018).
Löder, M. G. J., et al, *Focal plane array detector-based micro-Fourier-transform infrared imaging for the analysis of microplastics in environmental samples.* Environmental Chemistry, 563–581, (2015).

[9] Doublet, J., et al, *Distribution of C and N mineralization of a sludge compost within particle-size fractions.* Bioresource Technology, 1254–1262, (2010).
Primpke, S., et al, *an automated approach for microplastics analysis using the focal plane array (FPA) FTIR microscopy and image analysis.* Analytical Methods, 1499–1511, (2017).

NFSU JOURNAL OF LAW & ARTIFICIAL INTELLIGENCE
JLAI
Volume: 1, Issue: 1
July 2022
E-ISSN: 2584-1661

characterization methods such as the pyrolysis-gas chromatography-mass spectroscopy (py-GCMS) method, the analytical methods of vibrational spectroscopy (micro) (Fourier transform infrared (FTIR), μ-FTIR, Raman spectrophotometer, and others), and thermogravimetric analysis (TGA). Among all these analytical methods, vibrational spectroscopy using FTIR is the most practiced method of detecting microplastics in a sample. FTIR allows rapid identification of plastic particles while preserving the sample. As analysis results, the FTIR presents a single spectral profile containing characteristic peaks for each type of polymer contained in the sample. In some cases, during this analysis, the FTIR is also coupled with a microscope, the whole is called FTIR Microscopy (μ-FTIR) [10]. The μ-FTIR, for the detection of the different polymers contained in the sample, makes it possible to obtain hyperspectral images of high precision and spatial resolution showing the precise position of each polymer in the sample without any MPs/NPs size barrier. According to researchers, this method is more satisfactory than all the other methods. It makes it possible to automatically determine the chemical and morphological characteristics of continuous MPs/NPs in the sample by correlating images and spectra obtained with a reference library for pairing and assigning the spectra to simulate the similarity between the data. base and those of the sample in question [11]. Furthermore, the similarity results rely only on the characteristic numbers of each polymer contained in the reference library and which are arbitrarily assigned to the spectrum of the sample; which leads to errors in identifying the type of polymer. In addition, the similarity assessment is a labor-intensive process due to the size of the data, the diversity of plastic waste contained in the samples, and the presence of other organic pollutants which interfere in the process of identification and classification of the types of MPs and NPs present in the samples [12].

In recent days, through the use of artificial intelligence (AI)/machine learning, researchers have developed more automatic and robust computational methods to detect and classify MPs and NPs accurately (Figure 3), error-free and as fast as possible regardless of the size of the data. In this review paper the advanced detection and classification method of microplastics in environmental matrices using AI developed by researchers to perfect detections from FTIR/μ-FTIR.

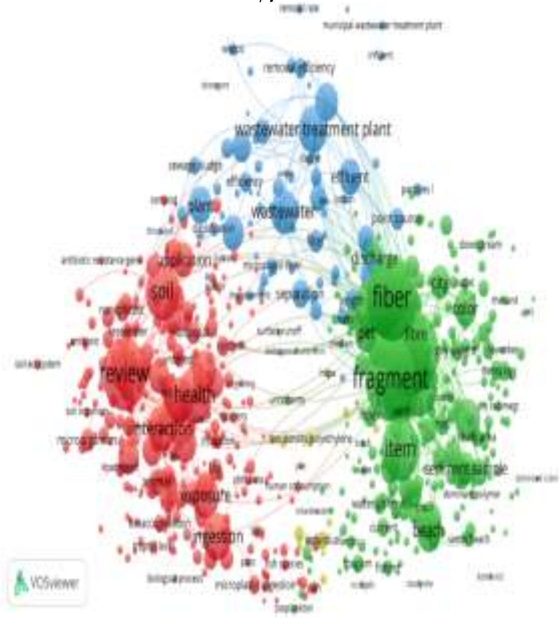

Figure 1: Overview recherche works on NPs/MPs and Domaine related

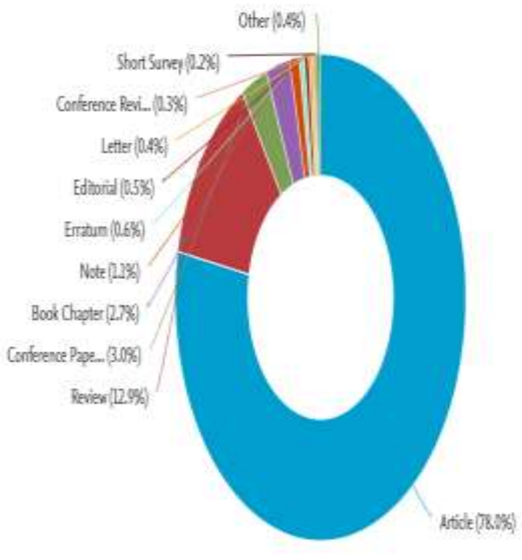

Figure 2: Type of document published related to MPs/NPs

[10] Kumar, R., et al, *Recent Developments in Extraction, Identification, and Quantification of Microplastics from Agricultural Soil and Groundwater*. Fate and Transport of Subsurface Pollutants, 125–143, (2021).

[11] Ali, I., et al, Micro- *and nano plastics in the environment: Occurrence, detection, characterization, and toxicity – A critical review*. Journal of Cleaner Production, 313, (2021). Also see Durante, C., et al, *A classification tool for N-way array based on SIMCA methodology.* Chemometrics and Intelligent Laboratory Systems,73–85, (2011). Hufnagl, B., et al, *Computer-Assisted Analysis of Microplastics in Environmental Samples Based on μFTIR Imaging in Combination with Machine Learning.* Environmental Science & Technology Letters,90–95, (2022).

[12] Prats-Montalbán, et al, *Multivariate image analysis: A review with applications.* Chemometrics and Intelligent Laboratory Systems, 1–23, (2011).

NFSU JOURNAL OF LAW & ARTIFICIAL INTELLIGENCE

JLAI
Volume: 1, Issue: 1
July 2022
E-ISSN: 2584-1661

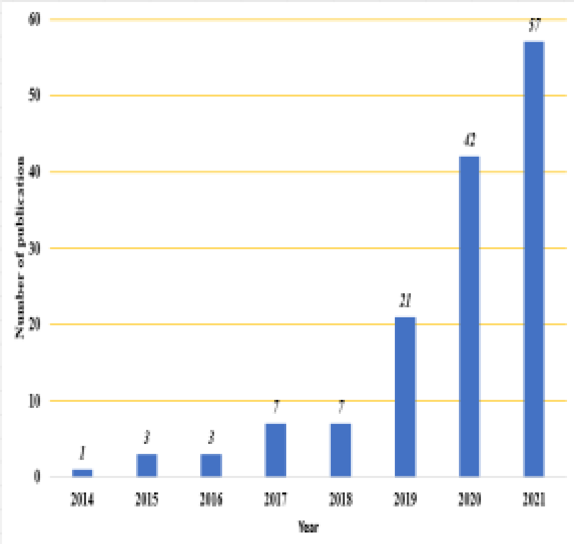

Figure 3: Number of publications on detection and identification methods of microplastics from samples

II. THEORETICAL OVERVIEW OF DIFFERENT MULTIVARIATE MODELS USED IN THE AI PROCESS

Artificial intelligence (AI) is the imitation of human intelligence by the use of a computer. It is a process that aims to create or transform ideas into a mathematical language called algorithms and execute them in a computer environment. AI is a computer innovation whose vision is to facilitate the most complex human tasks through the autonomous use of computers [13]. Today, AI popularly known as machine learning (ML), is used in several scientific fields such as engineering, medicine, the science of predictions, security, investigation, investigation of crimes, and others domaines. AI is a method of developing algorithms with the ability to autonomously and adequately process-specific data for good decision-making and/or succinct prediction[14]. In polymer detection and classification, AI uses supervised and unsupervised ML; For the supervised ML, the procedure consists to provide the algorithm with basic data called training data with class labels. After the identification of the different types of polymers present in any sample, apart from the algorithm through the computer, then follows the task of sorting and classifying each polymer into different classes. But by contradiction, this classification of polymer is done by the use of ML with the unsupervised algorithm: these are clustering algorithms. Referring to several kinds of research in the past, different types of algorithms have been developed for the identification and classification of MPs and NPs present in a sample. The following sub-sections focus on the overview of the algorithms used by AI in previous work [15]. Figure 4 is showing the differences between Supervised and Unsupervised Machine Learning.

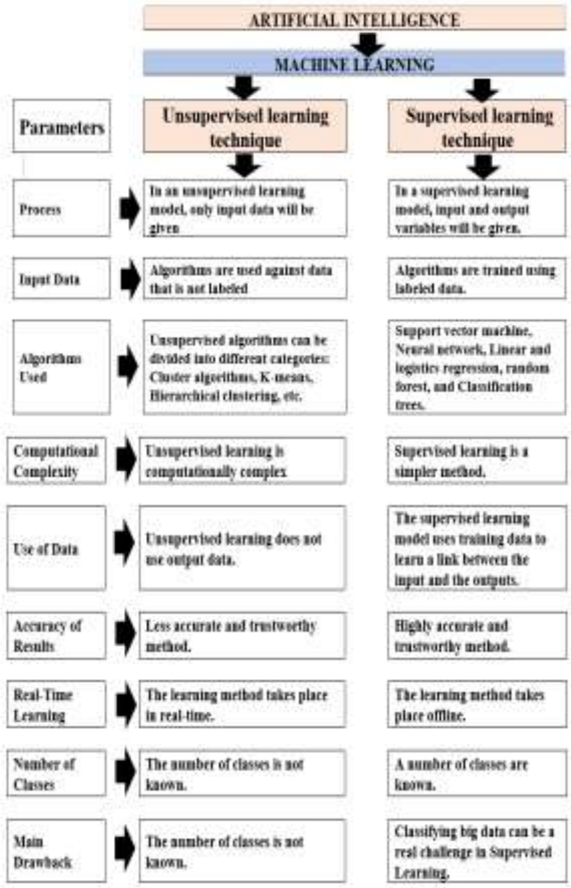

Figure 4: Difference Between Supervised and Unsupervised Machine Learning

[13] Post, C., et al, *Application of Laser-Induced, Deep UV Raman Spectroscopy and Artificial Intelligence in Real-Time Environmental Monitoring*, Solutions and First Results, 11, (2021).

[14] Comnea-Stancu, et al, *On the Identification of Rayon/Viscose as a Major Fraction of Microplastics in the Marine Environment: Discrimination between Natural and Manmade Cellulosic Fibers Using Fourier Transform Infrared Spectroscopy*. Applied Spectroscopy, 939–950, (2017).

[15] Cowger, W., et al, Critical *Review of Processing and Classification Techniques for Images and Spectra in Microplastic Research*. Applied Spectroscopy, 989–1010, (2020).
da Silva, V. H., et al, *Classification and Quantification of Microplastics (<100 μm) Using a Focal Plane Array–Fourier Transform Infrared Imaging System and Machine Learning*. Analytical Chemistry, 13724–13733, (2020)

JLAI
Volume: 1, Issue: 1
July 2022
E-ISSN: 2584-1661

III. AI: USED OF UNSUPERVISED ML

The main objective of unsupervised ML is to find different subsets of similar data in a data set. The given subsets detected in a data set are automatically labeled from the exploration algorithm. Clustering algorithms, without any human effort, group these subsets of data according to their common characteristics and properties [16].

For this purpose, other algorithms are also used for visualization in a simpler dimensional space. But their efficiency and accuracy are influenced by the distance metric chosen during each operation. For this, when detecting and classifying MPs and NPs, due to the complexity of the spectra and the size of the data, clear visualization of the subsets in a geometric space is often not possible [17].

Referring to previous work, five major algorithms have been used: Principal Component Analysis (PCA), Hierarchical cluster analysis (HCA), Hierarchical density-based spatial clustering of applications with noise clustering ((H)DBSCAN), K-means clustering, and Uniform manifold approximation and projection (UMAP).

1. K-means Clustering- In the case of K-means clustering, after the detection and classification of subsets in a dataset, the position of each data element in a cluster (subset) in the simple and optimized geometric space is determined by the distance K between spikes or elements in each different detect cluster. Referring to Wander et al., [18], if using spectral data collected from FTIR, the eight subset clusters of data detected projected by K-means clustering in geometric space are less detailed for visualization of the different types of and NPs contained in the sample dataset. The use of K-means clustering does not make clear and detailed discrimination between different types of polyesters [19].

2. Hierarchical Cluster Analysis (HCA)- The principle of HCA is to group each data element contained in a data set according to the resemblance criterion previously defined. This criterion of resemblance is expressed in a matrix of distances existing between each element two by two. The distance is null between two similar observations. In addition, observations with different resemblance criteria are expressed with a larger matrix distance. The HCA will then gather individuals iteratively to produce a dendrogram or classification tree. The classification produced by HCA starts from individual observation and henceforth is ascending. The HCA also produces hierarchical links producing larger and larger classes with subgroups embedded in each class. By cutting this tree at a certain chosen height, the desired partition will be produced [20].

Primpke et al., [21] report that manual clustering was superior to HCA in the case of the organization of spectral data obtained from FTIR from an analysis of detection of MPs and NPs present in a sample because of the similarity interferences that are created. with other database entries. These interferences are difficult to take into account by the algorithm. But he points out that the detailed data on the clusters is significant. Other research has evaluated the performance of HCA combined with principal component analysis. According to them, this combination makes it possible to check the accuracy of the classification results, especially for outliers [22]. The HCA made it possible to distinguish the natural cellulosic fibers from the artificial fibers[23].

---

[16] Ertel, W. *Machine Learning and Data Mining*, Introduction to Artificial Intelligence, 175–243 (2017). Also see Greener, et al, *A guIde to machine learning for biologists*. Nature Reviews Molecular Cell Biology, 40–55, (2022). Lussier, Fet al, *Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering*. TrAC Trends in Analytical Chemistry, 124, (2020).

[17] Ertel, W. *Machine Learning and Data Mining*, Introduction to Artificial Intelligence, 175–243 (2017). Also see Greener, et al, *A guide to machine learning for biologists*. Nature Reviews Molecular Cell Biology, 40–55, (2022).

[18] Wander, L., et al, *Exploratory analysis of hyperspectral FTIR data obtained from environmental microplastics samples*. Analytical Methods, 781–791, (2020)

[19] Basu, B., et al, *Development of Novel Classification Algorithms for Detection of Floating Plastic Debris in Coastal Waterbodies Using Multispectral Sentinel-2 Remote Sensing Imagery*. Remote Sensing, 13, (2021)

[20] Almeida, J. A. S., et al, *Improving hierarchical cluster analysis: A new method with outlier detection and automatic clustering*. Chemometrics and Intelligent Laboratory Systems, 208–217, (2007).

[21] Primpke, S., et al, *an automated approach for microplastics analysis using the focal plane array (FPA) FTIR microscopy and image analysis*. Analytical Methods, 1499–1511, (2017).

[22] Weisser, J., et al, *The Identification of microplastics based on vibrational spectroscopy data – A critical review of data analysis routines*. TrAC Trends in Analytical Chemistry, 148, (2022).

[23] Comnea-Stancu, Iet al, *On the Identification of Rayon/Viscose as a Major Fraction of Microplastics in the Marine Environment: Discrimination between Natural and Manmade Cellulosic Fibers Using Fourier Transform Infrared Spectroscopy*. Applied Spectroscopy, 939–950, (2017)

NFSU JOURNAL OF LAW & ARTIFICIAL INTELLIGENCE
JLAI
Volume: 1, Issue: 1
July 2022
E-ISSN: 2584-1661

IV. SPATIAL CLUSTERING OF APPLICATIONS BASED ON HIERARCHICAL DENSITY WITH NOISE GROUPING ((H)DBSCAN)

HDBSCAN is an algorithm for grouping into points, each data element contained in a data set according to the density of the cluster concerned. This grouping is prioritized based on the persistence of each cluster [24]. Wander et al.,[25], confirms that class discrimination is more detailed with HDBSCAN compared to that made by k-means clustering. HDBSCAN attribute has all the points representing each element in the clusters and merges each element in the cluster. According to Wander et al. [26], this merging of clusters could lead to an absurd number of clusters. In addition to your detection of particles of MPs/NPs contained from the spectral data, the absurdity of the clusters could lead to the non-consideration of a large number of spectra; So, so the application of HDBSCAN in the detection and classification of MPs and NPs is not too reassuring.

V. AN APPROACH BASED ON PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal component analysis (PCA) is one of the famous methods used by artificial intelligence in automatic exploratory analyses to extract features from a multivariate data set[27]. In the case of multispectral data coming from the identification of MPs/NPs from the FTIR, μFTIR, Raman spectrophotometer, and others, the PCA makes it possible to reduce the dimensions of the data in a new orthogonal space presenting in the maximum order of variance the different variables from the original data set[28]. This variance of different variables of the set obtained from the original data are linear combinations of the original variables: These are Principal Components (PC)[29]. At the end of this process, the various variables are projected into the orthogonal space as points called scores. These variables projected in space are grouped, are grouped by similarity or score, which allows seeing the differences in composition allowed by the components of the sample (Figure 4) [30] .In most cases of detection and classification of MPs and NPs present in a sample, the spectral data collected in the form of imaging from the μ-FTIR, are easily processed by PCA. The PCA made it possible to assign a class to each spectrum contained in the image and to prepare data of smaller size and easy to read [31]. The PCA alone is sufficient to process spectral data from a sample, classify the spectra of MPs/NPs, and facilitate their identification in the real world [32].

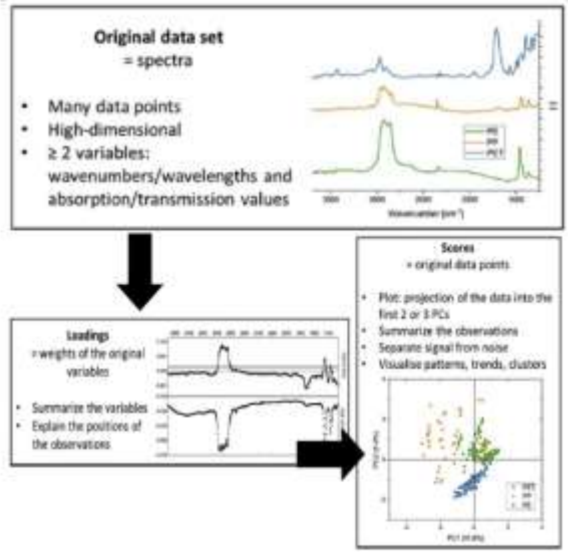

Figure 5: Illustration of the process of applying ACP in detection and classification from FTIR spectral data

VI. UNIFORM VARIETY APPROXIMATION AND PROJECTION (UMAP)

[24] Ankerst, M., et al, *OPTICS: Ordering points to Identify the clustering structure*. ACM SIGMOD Record, 49–60, (1999).
[25] Wander, L., et al, *Exploratory analysis of hyperspectral FTIR data obtained from environmental microplastics samples.* Analytical Methods, 781–791. (2020)
[26] Wander, L., et al, *Exploratory analysis of hyperspectral FTIR data obtained from environmental microplastics samples.* Analytical Methods, 781–791. (2020)
[27] Bro, R., & Smilde, *Principal component analysis*. Analytical Methods, 2812–2831, (2014).
[28] Granato, D., et al, *Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for a multivariate association between bioactive compounds and functional properties in foods: A critical perspective*. Trends in Food Science & Technology, 72, 83–90 (2018).
[29] Vitor, et al, *Classification and Quantification of Microplastics (<100 μm) Using a Focal Plane Array–Fourier Transform Infrared Imaging System and Machine Learning*, Analytical Chemistry, 4, (2020).
[30] Gautam, R., et al, *Review of multidimensional data processing approaches for Raman and infrared spectroscopy.* EPJ Techniques and Instrumentation, 1–38, (2015).
[31] Back, H. de M., et al, *Training and evaluating machine learning algorithms for ocean microplastics classification through vibrational spectroscopy*. Chemosphere, 287, (2022)
[32] Ballabio, D., et al, *Multivariate comparison of classification performance measures*. Chemometrics and Intelligent Laboratory Systems, 33–44, (2018).
Wander, L., et al, *Exploratory analysis of hyperspectral FTIR data obtained from environmental microplastics samples.* Analytical Methods, 781–791. (2020)

NFSU JOURNAL OF LAW & ARTIFICIAL INTELLIGENCE

JLAI
Volume: 1, Issue: 1
July 2022
E-ISSN: 2584-1661

UMAP is a similar approach to PCA [33]. UMAP also makes it possible to reduce the size of the spectral data. But UMAP, for this data dimension reduction, uses the nonlinear transformation method while maintaining the apparent structure of the data in a wider dimension space. This makes it possible to detect the most complex relationships between the points of each data element of the set. From the UMAP approach, the classification and identification of MPs and NPs more precise and real compared to PCA because it allows for differentiating the types of polymers, outliers, and the substrate of the filter used [34] .

VII. AI: USED SUPERVISED ML

Unlike unsupervised ML, supervised ML is autonomous seeding. For these types of models, the user must define the classes from the predetermined data to allow the algorithm to sort the data in the set into classes. After training the classes, the user should be able to re-classify points of dataset elements that do not fall on predefined sectors into the corresponding classes with high certainty and low bias[35].

VII. ARTIFICIAL NEURAL NETWORKS (ANN)

ANNs are computer systems whose mode of operation is similar to that of the nervous system of animals. ANNs, like the nervous system of animals, transfer information from neuron to neuron until the last one after which classes are generated[36]. They make it possible to decode the spectral image by cutting it into very small pixel blocks. According to Chaczko et al., (2018); Ertel, (2017b); Guo & Wang, (2021); Jiang et al., (2022), and Ng et al., (2020), and have an average accuracy of 78.5% detection and classification of polymers

VIII. AN APPROACH BASED ON SOFT INDEPENDENT MODELING OF CLASS ANALOGIES (SIMCA)

The SIMCA is a model similar to the PCA. Indeed, the SIMCA develops for each class category a PCA model and the limits within multivariate data. The SIMCA, by analogy based on mathematical processes, creates between the samples or the variables of a specific class and presents the various PCA models which define each category of class and the distance (equation 1) which separates them on both sides. other. The PCA models created and the distances calculated will allow the SIMCA model to define whether or not each sample belongs to a data category [37]. Since each of the samples will be tested to define its membership in many to one or zero class categories, there are limits to overlap. These overlapping limits cause a sample to belong to one or more and/or not to belong to a class category. As an advantage, the SIMCA model makes it easy to identify correlated samples from those that are not. In the case of the detection and classification of plastic types from multivariate data coming from the analysis of a sample from the FTIR, μ-FTIR, and Raman spectrophotometer, the updates of the basic data of the model are possible [38].

$$D_{i,g} = \sqrt{(T_{i,g}^2)^2 + (Q_{i,g}^2)^2} \qquad (1)$$

Where the T^2 is the Hotelling's T-Squared and Q is the matrices residual

IX. AN APPROACH BASED ON PARTIAL LEAST SQUARES

Partial Least Squares (PLS) regression is a mathematical precedent for creating smaller sets of uncorrelated components by reducing

[33] Gunnar Gruvaeus & Howard Waine, *Two Additions to Hierarchical Cluster Analysis†*, British Journal of Mathematical and Statistical Psychology, (1972). Also see Primpke, S., et al, *An automated approach for microplastics analysis using the focal plane array (FPA) FTIR microscopy and image analysis.* Analytical Methods, 1499–1511, (2017).

[34] Benedikt, H., et al, *Computer-Assisted Analysis of Microplastics in Environmental Samples Based on μFTIR Imaging in Combination with Machine Learning*, Environmental Science & Technology Letters, (2021, December 9).

[35] Post, C., et al, *Application of Laser-Induced, Deep UV Raman Spectroscopy and Artificial Intelligence in Real-Time Environmental Monitoring*, Solutions and First Results, 11, (2021).

[36] Baskin, I. I., et al, *Neural Networks in Building QSAR Models. In D. J. Livingstone (Ed.), Artificial Neural Networks,* Methods, and Applications, 133–154, (2009). Also see Burden, F., & Winkler, D. *Bayesian Regularization of Neural Networks. In D. J. Livingstone (Ed.), Artificial Neural Networks,* Methods and Applications, 23–42, (2019)

[37] Biancolillo, A., & Marini, F. *Chemometric Methods for Spectroscopy-Based Pharmaceutical Analysis.* Frontiers in Chemistry, 6, 576, (2018).

[38] Cocchi, M., et al, *Chapter Ten—Chemometric Methods for Classification and Feature Selection*, Comprehensive Analytical Chemistry, 265–299, (2018). Also see Muhamadali, H., et al, *RapId Detection and Quantification of Novel Psychoactive Substances (NPS) Using Raman Spectroscopy and Surface-Enhanced Raman Scattering.* Frontiers in Chemistry, 7, (2019). Vitor, et al, *Classification and Quantification of Microplastics (<100 μm) Using a Focal Plane Array–Fourier Transform Infrared Imaging System and Machine Learning*, Analytical Chemistry, 4, (2020).

predictors [39]. The least-squares are performed on the components by the regression and not on the initial data. In the detection and classification of MPs/NPs, the PLS consists in modeling the multivariate spectral data collected from the FTIR, μFTIR, and the Raman spectrophotometer by creating a correlation relationship between each variable of the set and the variables/parameter targeted or fixed [40]. This correlation created between the variables and the parameters set is based on the chemical/physical properties of the sample component and their chemical compositions [41]. Generally, the PLS regression model is a supervised model which reduces the dimensions or the size of the data to give rise to new variables which are called latent variables (LV). The latent variables (LV) are represented in a geometric space directed by the matrices of spectral data named matrix X and the matrix of parameters or fixed/targeted variables named matrix Y. the matrix Y, in the model, is a matrix that makes it possible to assign to each variable a corresponding membership class [42].

In practice, the application of the PLS model is based on the NIPALS algorithm. With the NIPALS algorithm, the matrices X and Y in two equations analogous to the PCA equation, where D and F are scores of each of the matrices, $\mathbf{P}^{\mathbf{T}}$, and $\mathbf{Q}^{\mathbf{T}}$ are the loading of matrices, $\mathbf{E}_1$ and E_2 are the matrices residuals [43].

$$\boldsymbol{X} = \mathbf{D}\boldsymbol{P}^{\boldsymbol{T}} + \boldsymbol{E}_{\mathbf{1}} \qquad (2)$$

$$\boldsymbol{Y} = \mathbf{D}\boldsymbol{Q}^{\boldsymbol{T}} + \boldsymbol{E}_{\mathbf{2}} \qquad (3)$$

For the assignment of the corresponding membership class to the variables of the matrix X from the matrix Y, the PLS model uses equation (4) to calculate the estimation probabilities (Z): W is the weight matrix; G is the residual matrix and b is the regression coefficient [44].

$$Z = \boldsymbol{T}\boldsymbol{Q}^{\boldsymbol{T}} + \boldsymbol{G} = \boldsymbol{X}\boldsymbol{W}(\boldsymbol{Q}^{\boldsymbol{T}}\boldsymbol{W})^{-\mathbf{1}}\boldsymbol{Q}^{\boldsymbol{T}} + \boldsymbol{G} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{G} \qquad (\mathbf{4})$$

With this membership class probability obtained, the model will use the discriminant method to assign each data variable of the sample a single corresponding class based on the percentage of the membership chance obtained from the model. The higher the percentage of probability of membership class concerning a fixed variable, the more the algorithm confirms the presence of this variable [45].

X. RANDOM DECISION FORESTS (RDF)

RDF is a tool for modeling and correlating the most complex data [46]. Referring to recent studies, the adaptation of RDF to the classification of MPs/NPs proves to be satisfactory. Hufnagl et al., (2022) confirm the sensitivity of RDF to detect and classify at nearly 94% performance from FTIR imaging. These results are confirmed by those of Hufnagl et al., (2022) and Kumar et al., (2021). After testing on images collected from μ-FTIR in drinking water, soil, sea salt, and sediment samples, the RDF can detect approximately 21 different classes of natural and synthetic polymers with 95.45% accuracy.

XI. FUTURE PERSPECTIVE AND CONCLUSION

It can be concluded that the use of AI is a very important and essential working tool in the detection and classification of MPs/NPs. All these algorithms presented by previous works prove useful in this process of size reduction of spectral data during the detection and classification of plastic particles present in the samples [47]. Moreover, at the moment, their

[39] Biancolillo, A., & Marini, F. *Chemometric Methods for Spectroscopy-Based Pharmaceutical Analysis.* Frontiers in Chemistry, 6, 576, (2018). Alos see VIdal, M., & Amigo, J. M. *Pre-processing of hyperspectral images. Essential steps before image analysis.* Chemometrics and Intelligent Laboratory Systems, 138–148, (2012).

[40] Gautam, R., et al, *Review of multIdimensional data processing approaches for Raman and infrared spectroscopy.* EPJ Techniques and Instrumentation, 1–38, (2015). Also see Muhamadali, et al, *RapId Detection and Quantification of Novel Psychoactive Substances (NPS) Using Raman Spectroscopy and Surface-Enhanced Raman Scattering.* Frontiers in Chemistry, 7, (2019). Also see Natalia, P. I. *Chemical Analysis of Microplastics and Nano plastics: Challenges, Advanced Methods, and Perspectives,* Chemical Reviews, (2008).

[41] Khan, I., Saeed, K., & Khan, I. *Nanoparticles: Properties, applications, and toxicities.* Arabian Journal of Chemistry, 908–931, (2019).

[42] Boulesteix, A.-L., et al, Partial *least squares: A versatile tool for the analysis of high-dimensional genomic data.* Briefings in Bioinformatics, 32–44, (2007).

[43] (Alin, 2009; Stott et al., 2017)

[44] Id

[45] De Luca, S., et al, *Class Modeling Techniques in Chemometrics: Theory and Applications.* In Encyclopedia of Analytical Chemistry, 1–24, (2018)

[46] B.N.Vinay Kumar, et al, *Analysis of microplastics of a broad size range in commercially important mussels by combining FTIR and Raman spectroscopy approaches,* ScienceDirect, (2021). Also see Hufnagl, B., et al, *Computer-Assisted Analysis of Microplastics in Environmental Samples Based on μFTIR Imaging in Combination with Machine Learning.* Environmental Science & Technology Letters, 90–95, (2022).

[47] Hufnagl, B., et al, *Computer-Assisted Analysis of Microplastics in Environmental Samples Based on μFTIR Imaging in Combination with Machine Learning.* Environmental Science & Technology Letters, 90–95, (2022). Also see Dong, M., et al, Automated *analysis of*

JLAI
Volume: 1, Issue: 1
July 2022
E-ISSN: 2584-1661

perfect and easy application is not yet level; future work should focus more on improving the algorithm models already proposed by the predecessors[48]. Also, evaluating the performance of the simultaneous combination of various algorithm models during detection and classification could be equally effective for MPs and NPs analysis [49].

microplastics based on vibrational spectroscopy: Are we measuring the same metrics? Analytical and Bioanalytical Chemistry. (2022).

[48] Back, H. de M., et al, *Training and evaluating machine learning algorithms for ocean microplastics classification through vibrational spectroscopy*. Chemosphere, 287, (2022).
Brandt, J., Mattsson, K., & Hassellöv, M. *Deep Learning for Reconstructing Low-Quality FTIR and Raman Spectra—A Case Study in Microplastic Analyses*. Analytical Chemistry, 16360–16368, (2021).

[49] Back, H. de M., et al, *Training and evaluating machine learning algorithms for ocean microplastics classification through vibrational spectroscopy*. Chemosphere, 287, (2022). Also see Brandt, J., Mattsson, K., & Hassellöv, M. *Deep Learning for Reconstructing Low-Quality FTIR and Raman Spectra—A Case Study in Microplastic Analyses*. Analytical Chemistry, 16360–16368, (2021). Also see Hufnagl, B., et al, *Computer-Assisted Analysis of Microplastics in Environmental Samples Based on μFTIR Imaging in Combination with Machine Learning.* Environmental Science & Technology Letters, 90–95, (2022).