

Article

A Handy Open-Source Application Based on Computer Vision and Machine Learning Algorithms to Count and Classify Microplastics

Carmine Massarelli * , Claudia Campanale  and Vito Felice Uricchio 

Water Research Institute, Italian National Research Council (IRSA-CNR), 70132 Bari, Italy; claudia.campanale@ba.irsa.cnr.it (C.C.); vito.uricchio@ba.irsa.cnr.it (V.F.U.)

* Correspondence: carmine.massarelli@ba.irsa.cnr.it

Abstract: Microplastics have recently been discovered as remarkable contaminants of all environmental matrices. Their quantification and characterisation require lengthy and laborious analytical procedures that make this aspect of microplastics research a critical issue. In light of this, in this work, we developed a Computer Vision and Machine-Learning-based system able to count and classify microplastics quickly and automatically in four morphology and size categories, avoiding manual steps. Firstly, an early machine learning algorithm was created to count and classify microplastics. Secondly, a supervised (k-nearest neighbours) and an unsupervised classification were developed to determine microplastic quantities and properties and discover hidden information. The machine learning algorithm showed promising results regarding the counting process and classification in sizes; it needs further improvements in visual class classification. Similarly, the supervised classification demonstrated satisfactory results with accuracy always greater than 0.9. On the other hand, the unsupervised classification discovered the probable underestimation of some microplastic shape categories due to the sampling methodology used, resulting in a useful tool for bringing out non-detectable information by traditional research approaches adopted in microplastic studies. In conclusion, the proposed application offers a reliable automated approach for microplastic quantification based on counts of particles captured in a picture, size distribution, and morphology, with considerable prospects in method standardisation.

Keywords: microplastics; computer vision; machine learning; automatic; quantification; classification



Citation: Massarelli, C.; Campanale, C.; Uricchio, V.F. A Handy Open-Source Application Based on Computer Vision and Machine Learning Algorithms to Count and Classify Microplastics. *Water* **2021**, *13*, 2104. <https://doi.org/10.3390/w13152104>

Academic Editor: Bing-Jie Ni

Received: 23 June 2021

Accepted: 29 July 2021

Published: 31 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Microplastics (MPs) are defined as “any synthetic or polymeric solid particles, with a regular or irregular shape, and with dimensions between 1 µm and 5 mm, of primary or secondary manufacturing origin, insoluble in water” [1]. Their primary origin refers to industrially produced materials (pellets and micro-beads) in the micro-size range used directly for specific applications (e.g., cosmetics, pharmacology, textile industry, medical diagnostics) or further processing. The secondary origin is related to the physical, chemical, or biological fragmentation of larger plastic debris (macroplastics) into smaller pieces [2–4].

In comparison to macroplastics with a larger size, the collection of microplastics from the environmental matrices and their relative isolation and quantification are complicated due to their small dimension and are considered as new challenges for analytical scientists; increasingly more sophisticated approaches and modern analytical instruments are required [2,5,6].

To compare results from different studies, it is necessary to define a standardised approach for sampling, extraction, and quantification of microplastic particles [7,8].

Experiments typically require accurate quantification of microplastics relying on counting each particle manually under the microscope. This is a rather complicated task because it results in a highly tedious process that requires a conspicuous demand in terms

of time and resources involved in collecting samples and processing them with thousands of particles to count per sample.

In light of this, in the present work, we propose an automatic approach to count the number of microplastic particles and classify them into different categories, laying the foundations for other promising MP research developments related to method standardisation.

The method proposed is simple and inexpensive, allowing MP morphology and size to be counted and analysed based on machine learning techniques that are gaining popularity due to automation and continuous availability. These techniques though need accurate hardware and an efficient computing model to achieve the desired success [9].

Recently, in the macroplastics field, remote sensing techniques coupled with machine learning approaches [10] have been adopted to quantify and classify beached macro litter, automatically resulting in a helpful tool for environmental pollution monitoring programs.

In microplastics research, only a few studies have tried to develop an automatic image analysis-based identification method to count and classify microplastic particles [9,11–13]. Some of them [14,15] used a ZooScan based system [16], in which digital images were post-processed with the Zooprocess and Plankton Identifier software, based on the ImageJ macro language [17,18], with the ability to attribute some morphological parameters to each object counted by the software including microplastics (e.g., the maximal distance between any two points along the boundary of the object, surface area). However, this application, initially developed to detect and classify zooplankton samples, applied to microplastics, implies the placement of every single item manually on the screen of ZooScan [15], providing a non-simultaneous identification of particles and a non-specific classification of them in morphology categories (e.g., pellets, fragments, fibres, etc.) already recognised in microplastics studies.

Therefore, the novelty of our work is using open-source Computer Vision and Machine Learning algorithms to count and classify microplastics by an environmental matrix, adopting a low-cost and straightforward approach. Moreover, our approach allows us to discover hidden information and drastically improves quantification and classification accuracy, reducing costs and analysis times.

The main objectives are the automatic and contemporaneous quantification of microplastic particles and their classification into four types of morphological interest and four size classes through machine learning algorithms and the implementation of a machine learning application in the MPs field.

2. Materials and Methods

2.1. Overall Operative Workflow

Instead of tedious and time-consuming manual particle counting and classification in size and shape categories, we developed a method based on the use of open-source computer vision (OpenCV) algorithms to extrapolate this information from the analysis of an acquired image [19]. In addition, based on the processing outputs of the OpenCV program library, we have written an early machine learning algorithm (principally based on if-then-else statements) with which we can count and classify the microplastics based on their different attributes, such as size ratio, morphology, perimeter, pixel intensity, and other parameters.

However, handwriting open-source programming code to apply these decision rules (better known as hand-coding) is characterised by some disadvantages because we applied them to obtain a decision only for a specific task in a single domain. For example, to be able to use the same code in other different applications, we would have to rewrite these rules, as we would not get any results; furthermore, if we noticed (based on operator experience) that another variable had to be introduced to obtain a better classification, we would have to redesign all the decision rules. Designing rules by hand requires a deep understanding of the problem, so it is very complicated in the MPs research sector due to a lack of knowledge. This is where Machine Learning (ML) comes in.

Currently, ML enters our lives daily, from protecting email to automatically tagging our friends in pictures or suggesting products and movies we like.

In recent years, the use of Deep Learning approaches for object classification [20–23] exhibited performance in complex tasks like never before [11].

ML is all about building mathematical models in order to understand data. ML enables computers to learn through experience to make predictions about the “future” using collected data from the “past” [24].

MP classification could be solved with a Deep Learning approach because learning models can adjust their internal parameters, modifying them to explain data better. With sufficient input data, we can ask the model to explain newly observed data.

In light of this, we developed an ML application to assess if there are more accurate classifiers for MPs identification.

We tested a supervised learning classification in which each data point is labelled or associated with a category or value of interest. At this point, we can apply supervised learning labelling of some example data, called training data, in order to make predictions about future data points (called test data). These standard procedures let us identify new particles in photos with the correct MP classes better and more rapidly.

On the other hand, to identify further information about MPs, hidden among large piles of data, we also used unsupervised classification (also known as data mining). One goal of an unsupervised learning algorithm is to organise the data somehow or to describe them. This can mean grouping them into clusters or finding different ways of looking at complex data to appear more straightforward.

The overall workflow adopted in the present work is reported here (Figure 1).

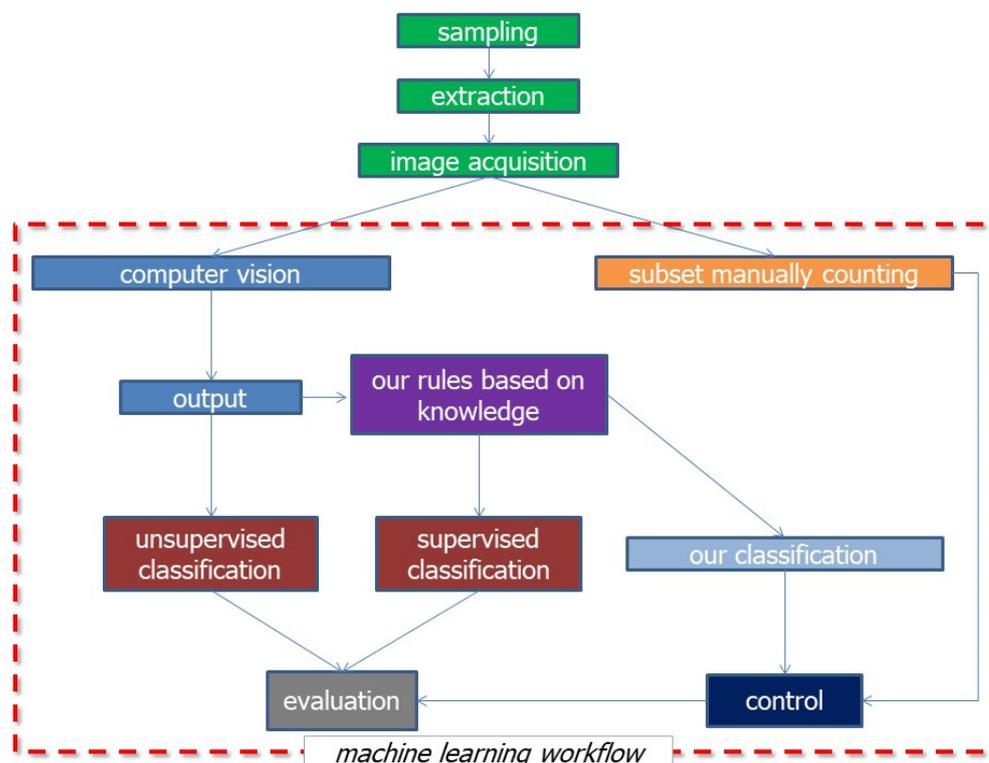


Figure 1. All the workflow adopted.

2.2. Sampling and Extraction of MPs

Freshwater samples were collected during a seasonal monitoring study of MPs performed in 2017–2018 in Ofanto river (South Italy) (41°17′20.22″ N; 16°06′92″ E - WGS84) [8].

Water surface samples (the first 45 cm of the river surface layer) were taken using three surface plankton nets (2.5 × 0.55 m) of 333 µm mesh size fixed in the middle of the river

simultaneously for two different time slots (11:00–13:00 and 13:00–15:00) for a total of six replicates collected for each campaign. The amount of net submerged was monitored and recorded, and an effort was made to maintain a consistent submersion depth throughout the sample duration.

At the end of the sampling step, the nets used were washed from outside to direct retained material (MPs and organic particles) into the collector tube. The samples were then transferred in labelled glass jars and stored at 4 °C until analysis. Firstly, wet sieving was carried out to retrieve the collected material, discarding the fraction of the sample larger than 5 mm. Wet peroxide oxidation (30% H₂O₂ and 0.05 M Fe(II)) followed by a density separation with NaCl was performed to digest labile organic fraction, leaving the MPs unharmed and isolating and extracting them.

Once all microplastic particles had been extracted, they were visually identified (Visual sorting) under a 40× digital microscope Keyence VH-Z 100 UR (Osaka, Japan); a further chemical identification of a subset of particles (about 3%) was also performed by Pyrolysis-Gas chromatography/Mass-Spectrometry (Py-GC/MS) to confirm their synthetic origin. Pyrolysis injector EGA/PY-3030D is from Frontier Lab (Koriyama, Japan), 7890B gas chromatograph and 7200 Q-ToF mass spectrometry are from Agilent Technologies (Santa Clara, CA, USA).

Further details regarding sampling strategy, processing, and characterisation methodologies were described in [6,8].

2.3. Image Acquisition

Once the samples had been processed in the laboratory to isolate and extract particles, a simple digital 12-megapixel smartphone camera was used to take pictures of them. MPs have to be distributed separately over a filter chosen based on their colour, increasing the contrast between particles and the background.

Image acquisition can be performed using any digital camera ranging from smartphone devices' simple camera to more sophisticated cameras of digital microscopes. It is better to use high-resolution cameras mounted on a tripod in such a way as to be able to acquire an image entirely orthogonal to the focal plane and, even better, if operated with a remote control to avoid any vibration that could produce a small shadow exchanged by a contour of the particle. Even the use of a smartphone camera has proved to be excellent because, to date, these devices have powerful software for managing the acquisition process, being capable reducing motion blur, with the same resolution, and automatically increasing the image quality (contrast, sharp colours, etc.). However, in hand-made photos, the possible presence of a minimum distortion does not affect the classification operated by the algorithm that considers several variables and not just one, as well as different rules and the different ratio between them, as shown in the next section.

The greater the camera's resolution power is, the lower the minimum size of the detectable microplastics will be.

The quality of images will result in a higher or lower level of resolution. High-resolution images imply a greater level of detail of particles but also a more prominent detection of any imperfection of the background. This may introduce some noise level that could be confused with some types of microplastics, resulting in the quantification of false positives. In Figure 2, an example of a picture obtained by a digital camera (Figure 2a) is reported, and the same picture is cropped (Figure 2b) to remove the useless contours that could generate errors in the processing, before applying the feature extraction.

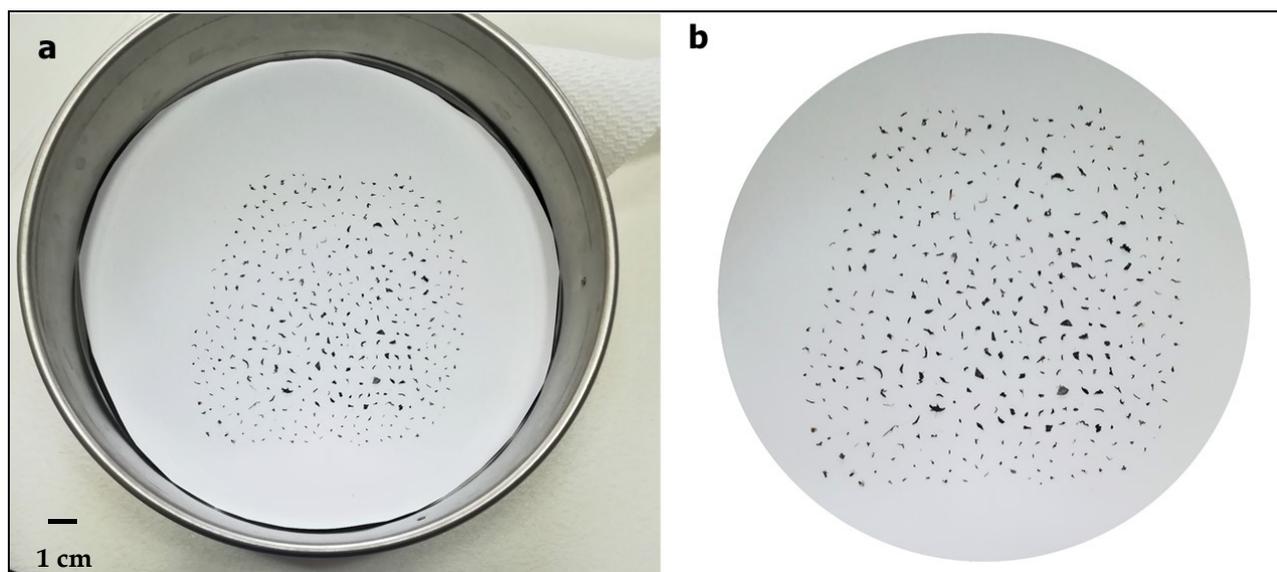


Figure 2. Image preprocessing steps: (a) Picture of an MPs sample in the sieve; (b) The same picture cropped.

2.4. Machine Learning Workflow

Computer vision research aims to convey the perceptions of humans to computers to make them able to sense the environment and take actions and learn from experience. Currently, computer vision systems are widely used to solve real-world problems, such as target recognition, manufacturing, photo interpretation, remote sensing, and navigation [25]. On the other hand, the field of machine learning is the idea that computer algorithms and Information and Communication Technologies (ICT) systems can improve their performance with time, evolving from general-purpose learning systems [26] to the symbolic learning of high-level knowledge [27] and artificial neural networks [28]. The trend of recent years is to consider machine learning algorithms as a highly robust tool to develop computer vision performance thanks to the learning-based approach. What is reported in this work is intended to be a further step in the complementary development of the two technologies and to be useful to solve a practical problem.

The developed method for our application is based on OpenCV Python Application Programming Interfaces (API). First, we wrote an algorithm based on OpenCV, an open-source BSD-licensed library of programming functions that access many open-source packages from the Scientific Python community, including several computer vision algorithms [19,29,30].

The OpenCV application is based, principally, on two processes.

The first process is called *feature engineering*, and it is principally based on a code implementation to find the best way to represent data—in this application coming from an image—that will be used in machine learning techniques.

The second is *feature extraction* which is implemented with Linear Blend Threshold, Binarisation, Bounding Box generation, extraction of particle features, and Classification based on size and morphology.

Linear Blend Threshold is the process of overlaying a foreground image with transparency (often the fourth channel of an image) over a background image. This transparency mask, called the alpha mask, is useful to the *Binarisation* in which we transform the sample image to black and white due to application of Otsu's thresholding [31].

The observation of a particular object in the images in Figure 3 (top left) suggests that, given the clear filter background, particles revealed are, in most cases, darker, such that by using the dark filter, the particles detected are the brightest.

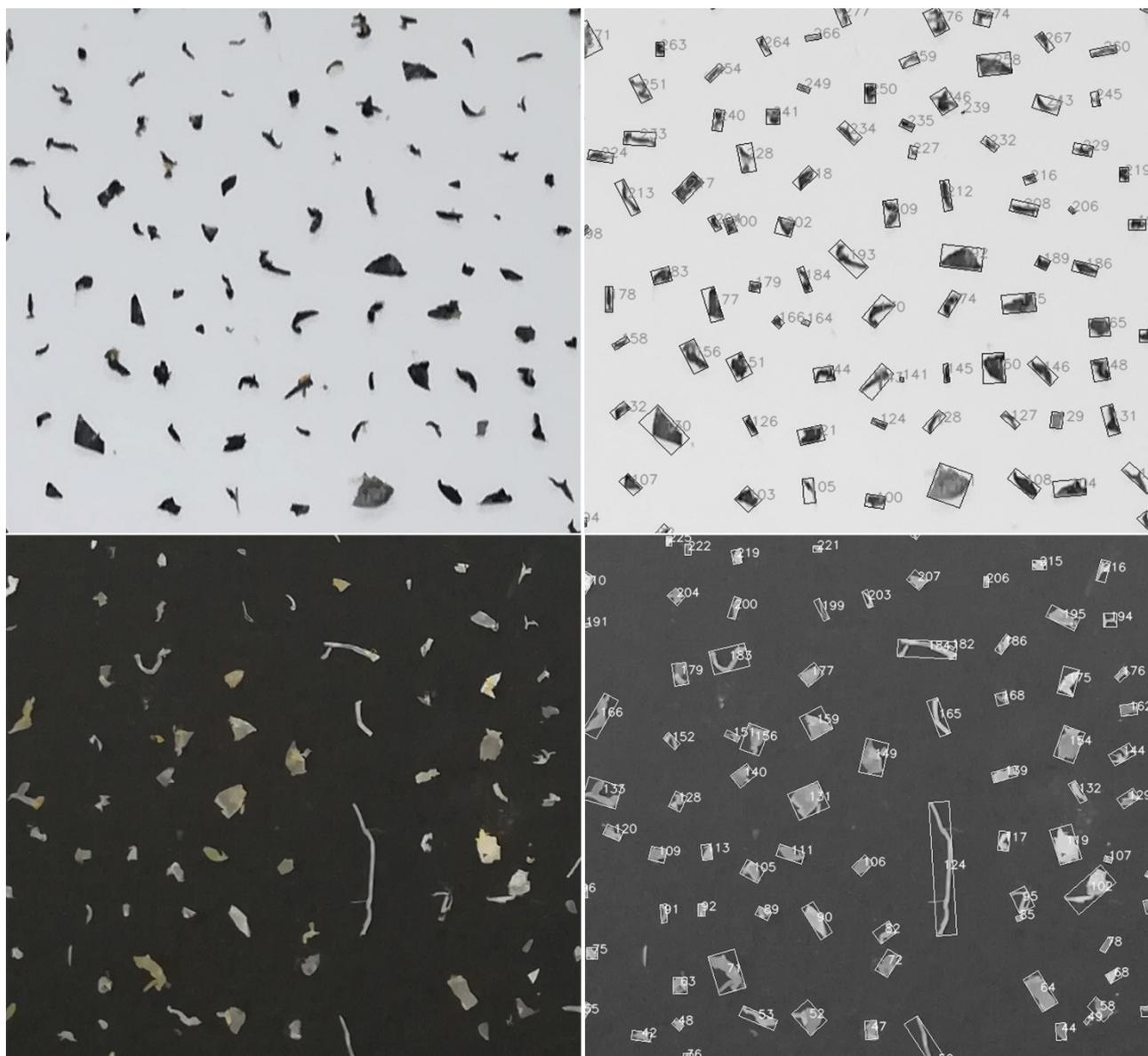


Figure 3. (left) MPs and background filters; (right) example of Bounding boxes generated with different background filters.

Before processing the image, we identified the threshold value to obtain the best possible results by contrasting the particles with the background as much as possible [11,12,32]. We considered, for the counting, all particles with a value higher than the threshold identified.

We can now find the different contours, such as area, perimeter, centroid, bounding box, etc. [33], and a certain weighted average of the image pixels' intensities. Simple properties of the image which are found via image moments include area (or total intensity) through Green's theorem [34,35], its centroid, and information about its orientation. We have as output much information and a drawn and numbered Bounding Box (Figure 3, right).

Subsequently, we tested our algorithm, an early machine learning system that uses hand-coded rules of if-then-else to process data through some rules implemented. We considered the ratio between the two main spatial dimensions that we called X and Y (the first is identified as the one that forms the smaller angle of incidence with the conventional X-axis, and vice versa for the second), the number of bounding box pixels and not null pixels, the mean value of pixel intensity, and others to classify MPs. Based on this, we obtained for each MP the following features:

- Perimeter of the particle (in mm);
- Area of the particle (mm²);
- Area of the particle bounding box (mm²);
- Ratio between the area of the particle and its bounding box;
- Y and X axes of the particle bounding box (mm);
- Centroids of Y and X axes;
- Number of pixels of the bounding box;
- Mean pixel intensity;
- Number of bounding box pixels;
- Number of not null pixels for each bounding box (equivalent to each particle's number of pixels).

After extracting the particle features, the next step was to output a class label for each particle in the image into one of the morphology and sizes classes defined as follows:

Morphology: The microplastics are observed in the environment in a wide variety of shapes. This information is a useful tool to indicate particles' potential origin [8]. The image analysis script differentiates particles according to their morphology as fragments, pellets, lines, and fibres. Based on the current typical microplastic morphology classifications [36–38] in this work, four classes of microplastics have been considered (Figure 4):

- Pellets.* This category corresponds to pre-production pellets, microbeads from personal care products and bead blasting, and other primary origin spheroids.
- Fragments.* Broken-down pieces of larger debris, such as plastic bottles.
- Lines.* Particles of fishing line and nets of longitudinal aspect with a thickness of about 1mm.
- Fibres.* Fibres from synthetic textiles of longitudinal aspect with a thickness <1 mm.

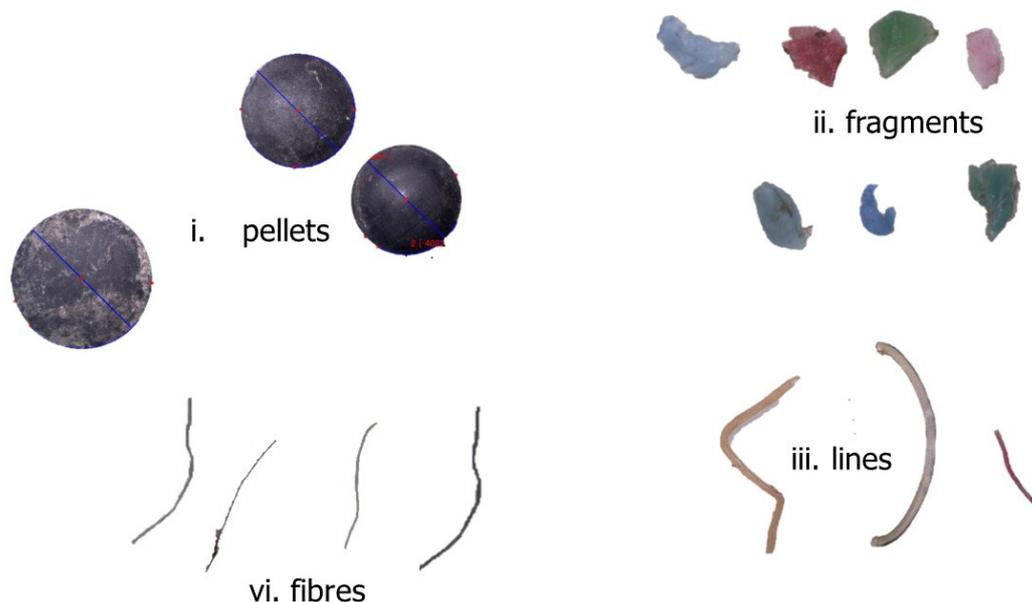


Figure 4. Types of microplastic morphologies identified for the automatic classification.

Below we report, for reasons of understanding, the rules used for the MPs classification in pseudocode format, which is a way of expressing an algorithm without conforming to specific syntax rules and is an efficient way to communicate ideas and concepts (note that indentation represents nested code).

```
START
```

```
for each particle in the image do this:
```

```
#here we consider a first geometric relationship of the shapes of MPs
```

```
if the ratio between the two dimensions of the bounding box is greater than 3.5:
```

```

    if mean pixel intensity is greater than 130.0 (on a scale of 0 to 255):
        assign classification 'fibre'
    else assign classification 'line'

#here we consider a second geometric relationship of the shapes of MPs
else if the ratio between numbers of not null pixels and of the product of the two
spatial dimensions of the particle is less than 0.4:
    if mean pixel intensity is greater than 130.0:
        assign classification 'fibre'
    else assign classification 'line'

else if the ratio between numbers of not null pixels and of the product of the two
spatial dimensions of the particle is greater than or equal to 0.4:
    if the ratio before calculated is greater than 0.7 and the ratio between the
two spatial dimensions of the particle is greater than 0.9 but less than 1.1:
        assign classification 'pellet'
    else (all the others cases) assign classification 'fragment'

END

```

The geometric parameters and the imposed rules derive mainly from the experience of the operators. The numbers entered are dimensionless as they are ratios. As you can see, referring to the concentration expressed previously concerning the specificity of algorithms written *ad hoc*, it lends itself to being highly efficient but only if applied to the reference domain for which it was created.

Size: As per the definition, the microplastics include all small plastic fragments <5 mm; the detection algorithm counts all objects from 5 mm and below.

Due to the diversity of sources, there exists a broad range of microplastics with variable shapes and sizes. These physical characteristics are important because they influence their distribution and impact on the environment. In particular, the MP's size is a critical aspect to be considered since the overlap between size categories of benthic and planktonic organisms and microplastics also enhances their potential ingestion by a wide variety of organisms. It is useful to distinguish at least two main size classes of MPs: <500 μm and 500 μm to 5 mm [39]. In the present work, four size classes have been considered:

- i. <500 μm ;
- ii. 500–1000 μm ;
- iii. 1000–2000 μm ;
- iv. 2000–5000 μm .

The size classification of particles using the ML algorithm has been performed for comparison with an object of known dimensions (5 \times 5 mm) drawn on the filter background where particles were placed. This known object was used as a reference to compute the size of the unknown particles. The algorithm was then able to count and divide particles into the size classes identified.

2.4.1. Supervised and Unsupervised Classification

Supervised Classification

Regarding the supervised classification, we used K-Nearest Neighbour (K-NN) algorithm to make a prediction of a new data point based on its neighbour. This is a non-parametric method used to classify data with discrete labels and regression and data with continuous labels [40,41]. K-NN algorithm needs a predefined number of training data closest in distance to the new point to predict the label from these. In K-NN, K is the number of nearest neighbours. The number of neighbours is the core deciding factor. K is generally an odd number. The choice of K depends on the characteristics of the data. Generally, as K increases, the noise that compromises the classification is reduced, but the class's choice criterion becomes more labile. Hence, we set K = 3, which provides a good compromise between an inaccurate classification and a labile and flat one [42,43].

Then, to apply the supervised classification, we proceeded to split data for two distinct phases:

1. *a training phase*, during which we aimed to train a machine learning model on a set of data that we called “the training dataset” (we stratify the training data with a specific function provided);
2. *a test phase*, during which we evaluated the learned (or finalised) machine learning model on a new set of never-before-seen data that we called “the test dataset”.

Different percentages of testing data to evaluate the response of K-NN were used. The distance chosen was based on the standard Euclidean distance implemented with Scikit-learn [44]. The algorithm was implemented with Python and machine learning application with OpenCV combined with Python, Scikit-learn, and Matplotlib.

To assess the validity and the goodness of the proposed classification and the performance of K-NN algorithm, we created a classification report, a text summary useful to understand in a more detailed way the classifier behaviour. The metrics are defined in terms of true and false positives and true and false negatives.

For each MP’s class, according to the percentage variation of the training dataset sample, we calculated accuracy, precision, recall, F1 score, and support described below.

Accuracy is a description of systematic errors, a measure of statistical bias (BS ISO 5725-1 1994), and describes the number of MPs classified correctly towards the total number of particles in our test set. Precision represents a classifier’s ability not to label an instance positive that is negative (for each class, it is the ratio of true positive to the sum of true and false positives); in other words, for all those classified positives, it represents the correct percentage. Recall is the ability to find all positive instances for each class and is the ratio of true positives to the sum of true positives and false negatives; in other words, for all those that were positive, it represents the correct percentage classified correctly. F1 score is a weighted harmonic mean of precision and recall and takes values from 1.0 (best) to 0.0 (worst). As a rule of thumb, the weighted average of F1 should be used to compare classifier models and not global accuracy [45,46]. Support is the number of actual occurrences of the class in the specified dataset. Imbalanced support in the training data may indicate structural weaknesses in the classifier’s reported scores and could indicate the need for stratified sampling or rebalancing.

Unsupervised Classification

Regarding the unsupervised classification, we used k-means clustering which searches for a predetermined number of k clusters (or groups) within an unlabelled multidimensional dataset.

It is based on two simple assumptions [24]:

1. the centre of each cluster is simply the arithmetic mean of all the points belonging to the cluster;
2. each point in the cluster is closer to its centre than to other cluster centres.

Similarity measures are based on Euclidean distance. Our goal is to identify the intrinsic properties of data points that make them belong to the same subgroup. We set the number of clusters to 4, known beforehand.

2.4.2. Control Tests (Subset Manually Counting)

Furthermore, to assess the proposed early ML algorithm’s validity, we manually counted the particles for five selected images corresponding to five different samples, based on our experience. We classified them in the different classes previously chosen (fragments, pellets, lines, fibres).

In addition, a simple software (tentatively called *Cont@tutto*) with a graphical user interface was created to facilitate the manual counting by a click. A background grid allows for better detection and counting of particles. Furthermore, with the mouse’s click, it is possible to mark (with a red dot) any desired particle (Figure 5).

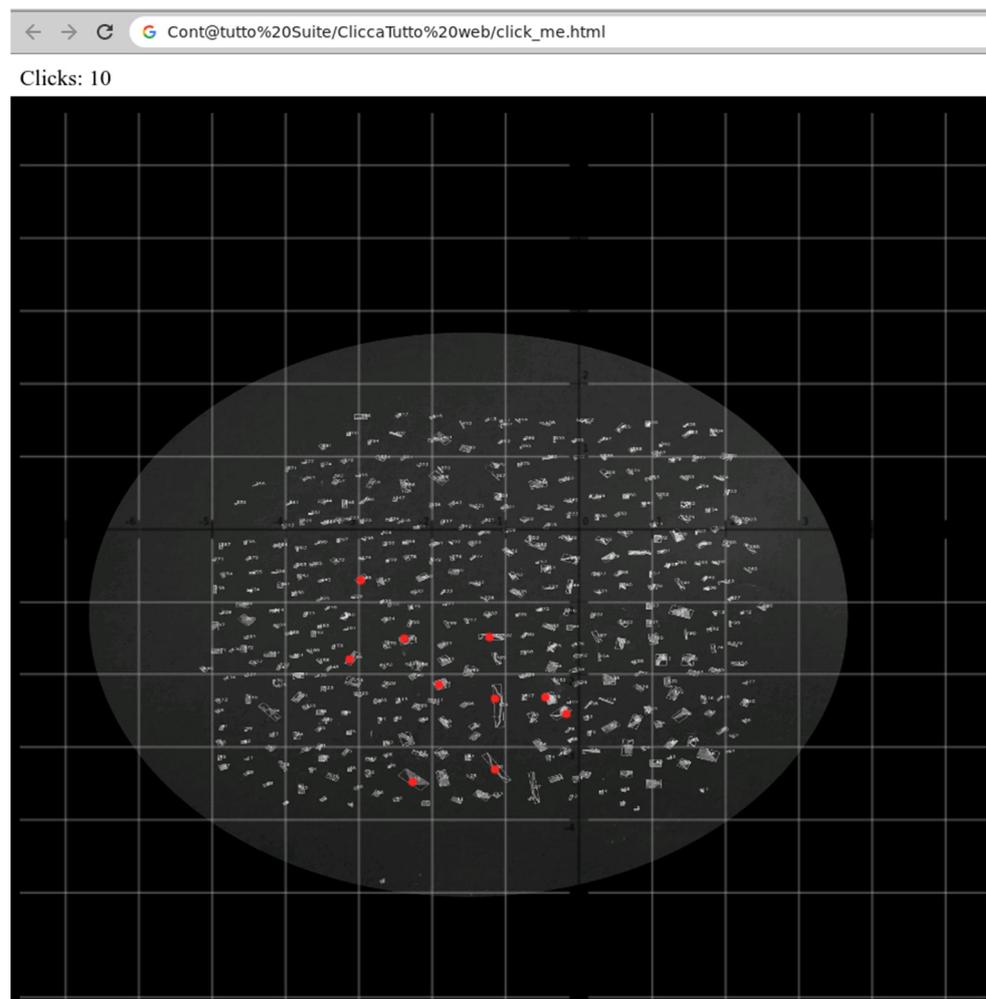


Figure 5. Details of graphical user interfaces used for manual counting.

3. Results

3.1. Manually Counting vs. Early Machine Learning

In order to verify the reliability of the results provided by our early ML algorithm and to test the performance and the validity of the proposed methodology for automatic counting and classification of particles, we compared these results to the manual counting of particles under the microscope by a human expert operator. These data represented the ground truth and were also used as training data for the supervised application and to compare the results of both classifications (supervised and unsupervised).

For the experimental setup, the particles were obtained from five images corresponding to five different samples. The samples, manually counted, contained 2501 particles subdivided into 2261 fragments, 51 pellets, 125 lines, and 64 fibres (Table 1).

Table 1. Manual vs. automatic counting of particles.

Sample	Manual Counting	Automatic Counting	Dev st. %
1	416	417	0.7
2	398	395	2.1
3	384	377	4.9
4	505	505	0.0
5	798	800	1.4
Total	2501	2494	4.9

As previously described, the results obtained by an expert (manual) and results revealed by the ML algorithm (automatically) were compared.

It can be observed that the counting results between the two sets of data (Table 1) are very similar with a standard deviation ranging from 0 to 4.9% among the samples with a total error of 4.9% in the total number of detected particles. Depending on the sample, the algorithm sometimes overestimated the real number of manually counted particles, and at other times underestimated them. The overestimation could happen if the picture had a very high resolution, and sometimes a dust grain could be confused with a particle; vice versa, underestimation could occur when the resolution was low, and the particles were too close each other.

However, these differences fall within the acceptable margin for monitoring studies [47,48].

Figure 6 represents the comparison between the manual and automatic classification of the number of total particles subdivided in the different morphology categories (expressed as the sum of five samples) and the separated results for each sample and class (Table 2).

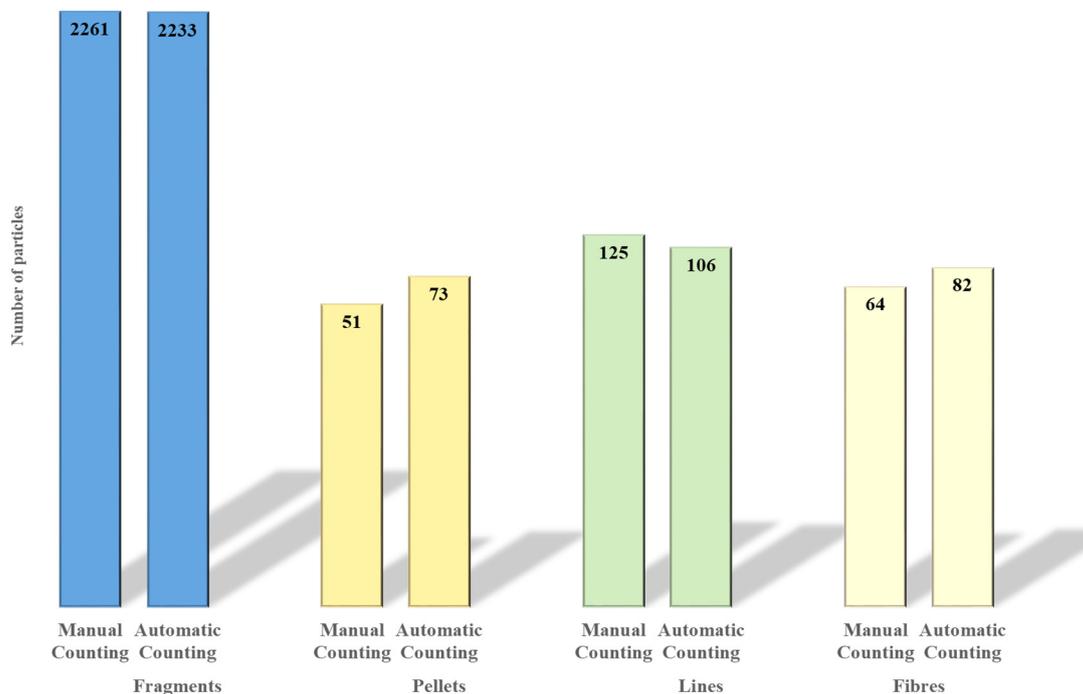


Figure 6. Manual vs. automatic classification results. Results are expressed as the sum of the different shape category in the five different samples.

Table 2. Manual vs. automatic classification (early machine learning algorithm) results per class.

Sample n.	Fragment			Pellet			Line			Fibre		
	Manual Counting	Automatic Counting	Dev st. %	Manual Counting	Automatic Counting	Dev st. %	Manual Counting	Automatic Counting	Dev st. %	Manual Counting	Automatic Counting	Dev st. %
1	382	372	12.7	9	12	2.1	4	12	5.7	21	29	5.7
2	348	346	1.4	5	9	2.8	10	0	7.1	35	40	3.5
3	337	333	2.8	5	8	2.1	42	26	11.3	0	2	1.4
4	462	454	5.7	6	12	4.2	37	38	0.7	0	0	0.0
5	732	728	2.8	26	32	4.2	32	29	2.1	8	11	2.1
Total	2261	2233	19.8	51	73	15.6	125	106	13.4	64	82	12.7

The fragments class obtains the worst results.

The shape categories’ differences are quite similar, showing a maximum value of the standard deviation of 19.8% (for fragments) and a minimum of 12.7% (for fibres).

Being three-dimensional particles, they acquire different conformations in space once positioned on a flat surface. For example, coiled lines are misclassified as fragments as they

lose their characterising features such as the “Ratio between the area of the particle and its bounding box”, helpful for their classification. As well as rolled fragments are also often erroneously classified as pellets because so arranged, they take the form of a pellet.

To evaluate the reliability of the results related to the automatic size classification of microplastics provided by our ML algorithm, a comparison with a manual measure of each single particles under the microscope would have been very time-consuming and unrealistic. In microplastic studies, the evaluation of size particles is usually performed using image processing software generally coupled to that those of microscopes. Taking the measures of each particle singularly would be unthinkable. Therefore, we chose to calculate some statistics of our known object that we drew on the background and used as a reference to calculate the unknown particles’ dimensions. In each xls sample report, the sizes of the known object (expressed as X and Y axes of the bounding box surrounding the object) were always reported and used to distribute particles in the previously defined size classes. We considered a set of 28 samples, and we computed the measure of the amount of variation (standard deviation), mean, median, maximum, and minimum values of the sizes of the known object as parameters to test the algorithm’s performance. Statistics reported in Table 3 show a mean value of X and Y of 5.08 ± 0.3 and 5.06 ± 0.3 , respectively; a maximum value for X and Y of 5.79 and 5.62; and a minimum value of 4.64 and 4.52.

Table 3. Sizes of X and Y calculated by the ML algorithm for an object of known dimension (5×5 mm) for a subset of 28 samples. At the end of the table, some statistics calculated for the subset are present.

Sample	Sizes of a Known Object		
	n.	X (mm)	Y (mm)
1		4.89	5.07
2		4.67	4.78
3		5.31	5.00
4		5.02	4.78
5		5.21	5.27
6		4.72	4.52
7		4.93	5.02
8		5.31	5.49
9		5.07	5.32
10		4.89	4.75
11		5.02	5.36
12		4.92	4.92
13		4.78	4.89
14		4.91	5.11
15		5.01	5.04
16		5.11	4.98
17		4.92	4.55
18		5.21	5.30
19		4.98	5.17
20		4.64	4.75
21		4.78	4.89
22		5.07	4.79
23		5.24	5.11
24		5.79	5.55
25		5.05	5.46
26		5.79	5.45
27		5.53	5.45
28		5.45	5.01
Mean		5.08	5.06
Median		5.02	5.03
St. dev.		0.30	0.30
Max Value		5.79	5.62
Min Value		4.64	4.52

3.2. Supervised vs. Unsupervised Classification

3.2.1. Supervised Classification

Concerning the supervised classification, we identified the best parameters that we thought could help classify MPs' different morphologies using a supervised approach. As we mentioned before, in the supervised learning approach, each data point is associated or labelled to a specific category of belonging. The algorithm will use this information to predict the category to which the MPs, not yet labelled, belong (this is called a classification problem to solve with a machine learning approach). These parameters were used as input data for the K-NN algorithm using a random training dataset of 50%.

In Figure 7, the calculated classification (by a human expert) and supervised classification (by the K-NN algorithm) are compared. The figure's left column shows the results of the classification operated by our classification algorithm, while the right column shows the results of the classification resolved with the machine learning approach. Different parameters deemed most interest are plotted on the X and Y axes. The classification is obtained plotting "X-axis of the particle bounding box vs. Y-axis of the particle bounding box" (Figure 7a,b), "area of the particle vs. mean pixel intensity" (Figure 7c,d), and "numbers of bounding box pixels vs. not null pixels for each bounding box" (Figure 7e,f).

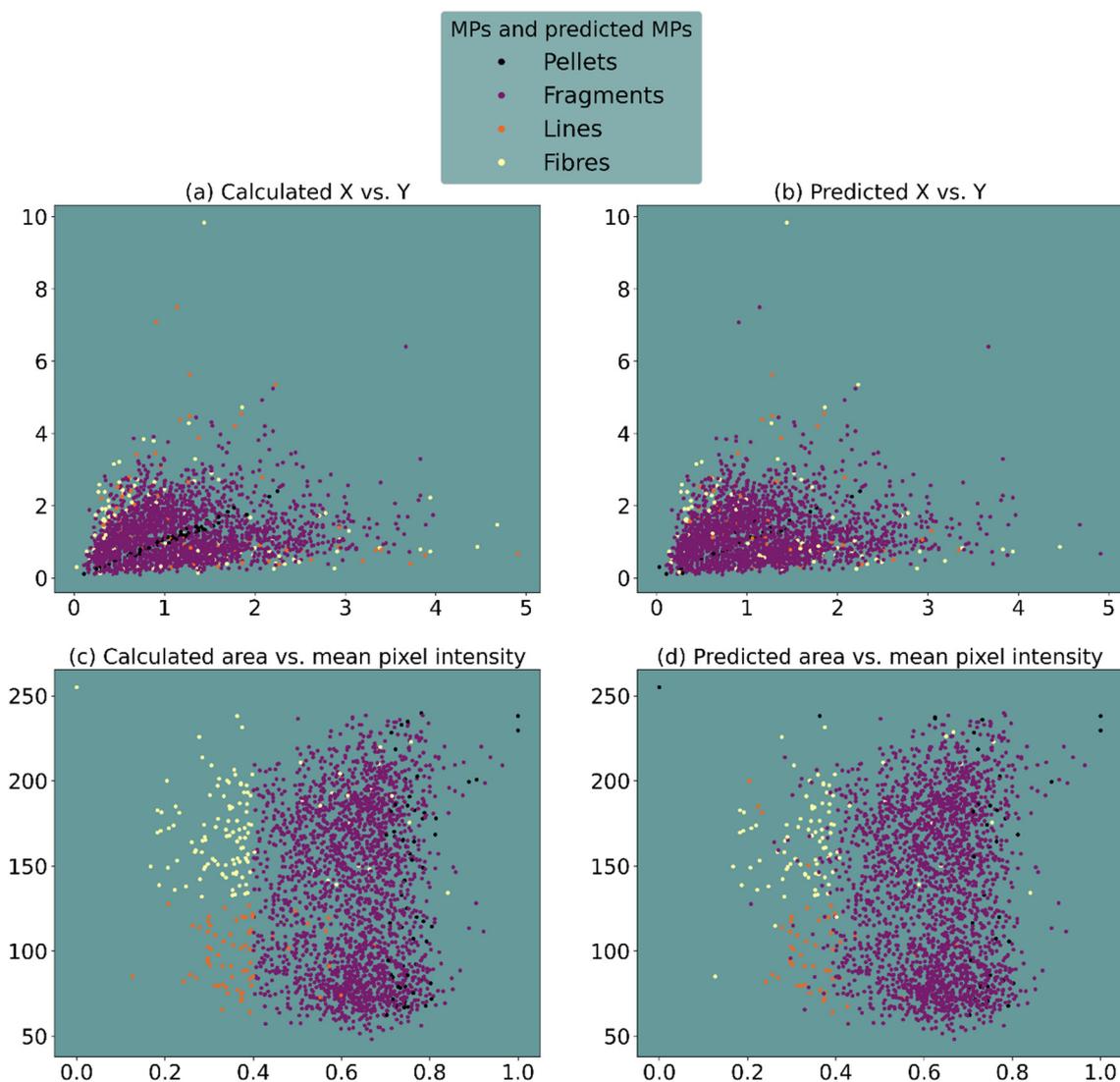


Figure 7. Cont.

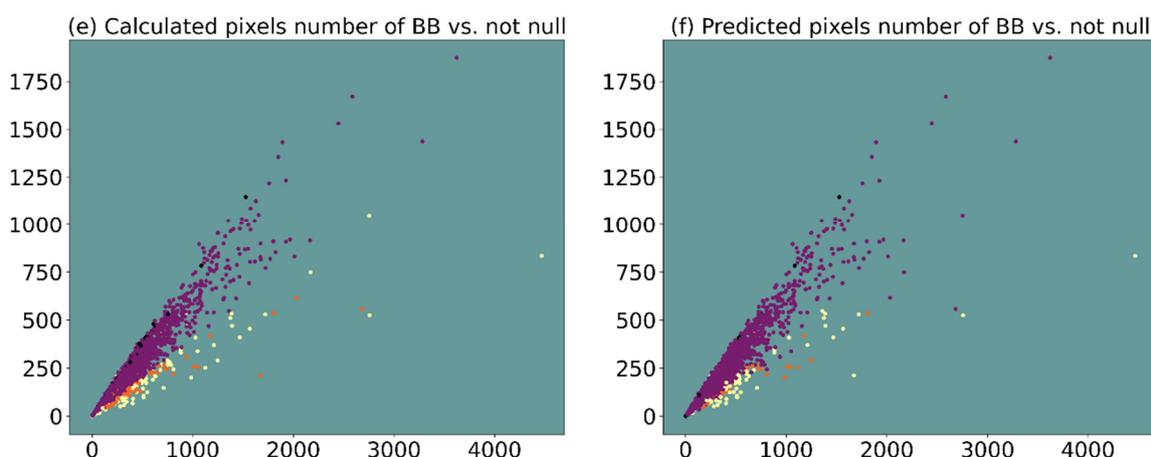


Figure 7. Plotted results of supervised classification. The “Calculated” column on the left (a,c,e) refers to data from a human expert, while the “Predicted” column (b,d,f) refers to values from K-NN algorithm. K-NN results plotted data were obtained using a training dataset at 50%.

Looking at Figure 7, we can observe the presence of some clusters.

In the classification obtained by a human expert (in the left column of Figure 7), in the subplot (a), it is clear that the pellets (purple dots) are arranged in a well-defined region as if they divided the plot into two parts. This is in line with their highly symmetrical spherical morphology and their well-strict primary origin [49,50]. Conversely, fibres (yellow dots) arrange themselves at the ends of the plot, leaving the central part to all the fragments (much more common).

In the subplots Figure 7c,d three clusters, although not perfectly symmetrical, are visible: the fibres and lines (green dots) have a visible distribution region while fragments and pellets overlap.

In subplots (e) and (f), all four classes of particles are evident which alternate in the spatial distribution in a radial pattern divided principally into two groups: fragments and pellets and fibres and lines.

In the classification obtained by the K-NN algorithm (in the right column of Figure 7), subplots (b) and (f) result as the best correspondence to reality, although in the former, the isolation of pellets (purple dots) is almost lost, and in the latter, there is the identification of lines (green dots). In the subplot (d), the measures do not indicate well-defined clusters, and therefore, there seems to be no clear separation between MP classes.

The performance of K-NN has been evaluated calculating some classification metrics (accuracy, precision, recall, F1 score, and support), presented in Table 4.

In line with the theory, the results highlight an increasing accuracy value parallel to enhancing the training data with the best value (0.922) obtained for a percentage of training data of 70% (Table 4). The precision, recall, and F1 score show values higher than 0.90 for the fragments, indicating the good performance of the K-NN classifier for this class, even using low rates of training data. On the other hand, for lines and fibres classes, acceptable results of the same measures are achieved only by using a high percentage of training data (70%).

The pellets are difficult to identify as they are often overestimated wildly when the precision and recall values are <0.01 . To better identify *pellets*, it seems necessary to set a training data value of 50%; in fact, they have entirely different origins and could be considered real outliers compared to other MPs. Indeed, increasing the number of training data to 70%, the classifier still has difficulty, identifying other fragments belonging to this class.

Table 4. Performance of K-NN algorithm using different percentages of training data.

Training Data	Accuracy	Typology	Precision	Recall	F1 score	Support
10%	0.905	Pellets	<0.01	<0.01	<0.01	60
		Fragments	0.91	0.99	0.95	1990
		Lines	0.67	0.17	0.27	60
		Fibres	0.63	0.22	0.33	108
20%	0.906	Pellets	<0.01	<0.01	<0.01	53
		Fragments	0.91	0.99	0.95	1769
		Lines	0.74	0.26	0.38	54
		Fibres	0.79	0.23	0.35	96
50%	0.916	Pellets	0.14	0.03	0.05	33
		Fragments	0.93	0.99	0.96	1105
		Lines	0.59	0.29	0.39	34
		Fibres	0.77	0.4	0.53	60
70%	0.922	Pellets	<0.01	<0.01	<0.01	20
		Fragments	0.93	0.99	0.96	664
		Lines	0.75	0.45	0.56	20
		Fibres	0.78	0.5	0.61	36

3.2.2. Unsupervised Classification

Looking at the results obtained from the unsupervised classification (column right of Figure 8) compared to those of the classification provided by the human expert (left column left of Figure 8), we can observe that the algorithm tends to classify the particles in a very different way.

In subplot Figure 8b, two distinct classes (purple and yellow dots) are evident, and an overlap, especially between classes 2 and 3, is observed. In the subplot Figure 8d, it is not possible to observe distinct patterns, but all the four classes are mixed, as already partially observed in the supervised classification. Therefore, the areas and pixel intensity parameters do not lend themselves at all to an unsupervised classification.

On the other hand, in the subplots Figure 8f, anisotropic clusters tend to form because the unsupervised algorithm identifies the component of the ratio along the diagonal of the graph as a discriminant among classes, and three different classes are well isolated (purple, yellow, and blue dots).

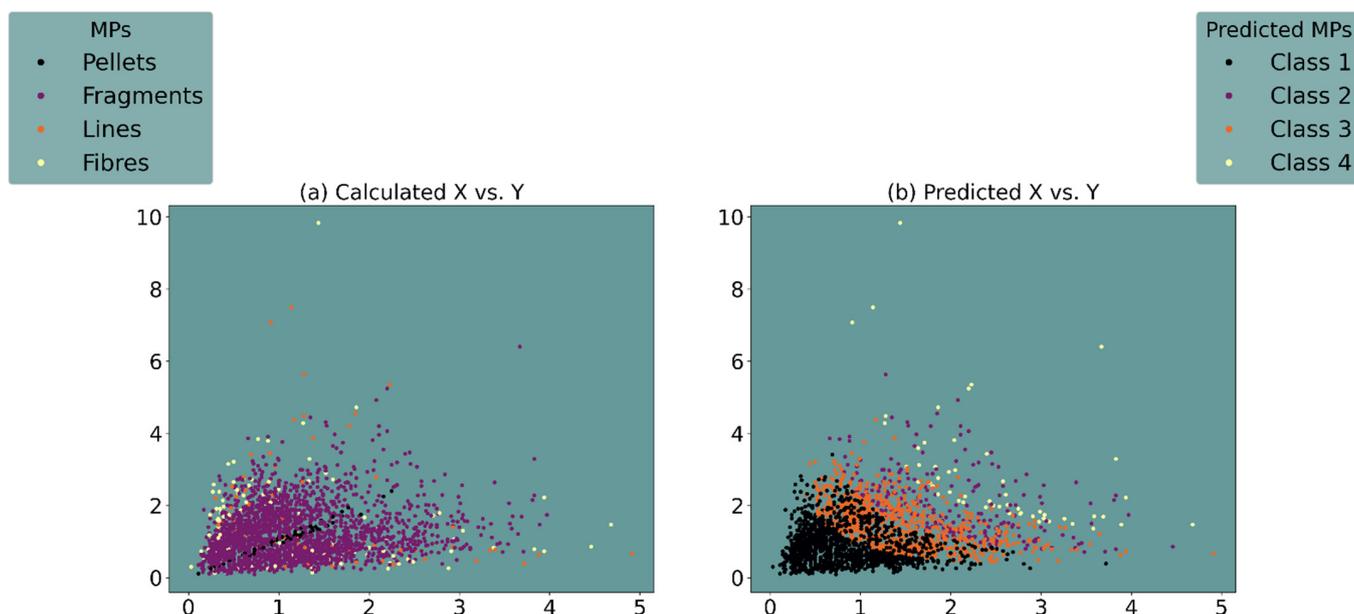


Figure 8. Cont.

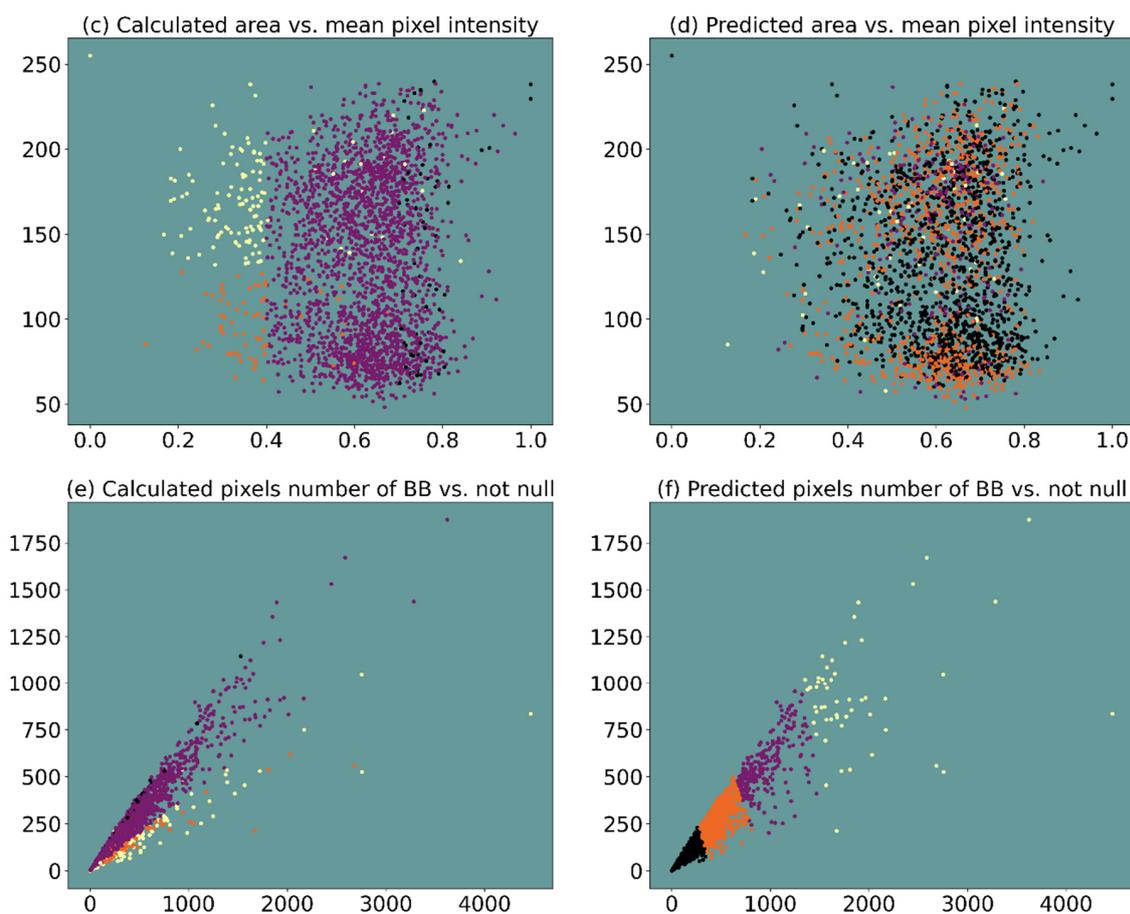


Figure 8. Plotted results of unsupervised classification. The “Calculated” column on the left (a,c,e) refers to output data from a human expert, while the “Predicted” column (b,d,f) refers to values from unsupervised classification.

This indicates that variations not in the ratio between the parameters but between the numerical values are preferred for this dimensional classification, in line with the subplot (b) (the same thing happens by providing the normalised or standardised values as input data).

4. Discussion

Overall, the MPs’ automatic counting and classification through early ML algorithms provided promising results substantially comparable to the human expert’s outcomes. According to the achieved results, the differences revealed fall within the acceptable margin of error for monitoring studies [47,48] that try to establish plastic debris’ evolution over time rather than obtain a snapshot of a single moment.

The most satisfactory results have been achieved concerning the particle counting process. It is good to underline that human errors are avoided by automating this phase, which can arise due to the tiredness caused by the long processing times of the images needed in the manual particles count.

The automatic classification of particles in different shape categories achieved good results, too, even if sometimes misleading classifications of particles occurred. This aspect certainly has to be improved, including other features and new rules for classifications of particles, but it is an optimum start point. Furthermore, once the process is better, it will provide a standardised method to classify microplastics based on objective parameters, without leaving room for subjective evaluations.

Last but not least, an aspect to consider is the considerable advantage of this automated methodology concerning saving time.

Once particles are positioned on a base with an appropriate background, simply take a picture. A report will be available in a few seconds with all the information about particles present in the photo with their relative classification in morphological and dimensional categories and the total count. However, it is worth specifying that to achieve satisfactory results, care has to be taken to accurately place particles over the background, separating them one by one manually to avoid overlapping. This process indeed requires time, but it is an operation that should also be done in the classical counting process of particles under a microscope. Moreover, operating in the traditional method, by moving the microscope's slide on which the particles are positioned to focus them, they will move again, thus implying a further waste of time. With our automated methodology, on the other hand, this operation must be performed only initially, and once the image has been acquired, we work in post-processing.

Comparing the results provided by a human expert with those of the K-NN algorithm (Figure 7), it is possible to observe the tested classifier's good performance. This is confirmed by the calculated value of accuracy that always showed values higher than 0.9, even using low percentages (10%) of training data. Logically increasing the rate of training data, the accuracy also enhances. The obtained results are quite promising, also comparing our data to [12,51], for whom, when evaluating different algorithms to count and classify MPs, the K-NN exhibited the worst performance with an accuracy of 0.721. Even if other authors [20,52,53] have recently started to use machine learning approaches in the MPs field, none of them has the same purposes, and therefore, we cannot compare our achieved performances. On the other hand, concerning the goodness of the application for the classification of MPs in different classes (fragments, pellets, lines, and fibres), we can observe that the morphology best identified is that of the fragments followed by fibres and lines, using high percentages of training data as confirmed by the precision, recall, and F1 score measures. This result can be easily explained as a suboptimal balance of the different types of MPs in the samples taken in the field and used as a starting dataset with a prevalence of fragments. However, our data came from a real study case.

Analysing the results related to the unsupervised classification (Figure 8), it is evident that clusters of variable geometries emerge in the "predicted" column and that the algorithm tends to classify the particles in a very different way from the "calculated".

The observation of subplots Figure 8b,f suggests that the unsupervised classification could probably reveal hidden information. Class 1 (purple dots) could be associated with the fibres class due to their smaller sizes with respect to the other MP classes, which causes them to occupy the positions of the corners in the two graphs and indicates their well-known prevalence in the environment [54–56], which determines their numerical abundance in the graphs. The underestimation of this class of MPs in the real situation (calculated) is most likely due to the sampling methodology adopted. Indeed, we used traditional methods to collect microplastics using 333 μm mesh nets that are too coarse to sample most textile fibres, thus failing to sample them adequately. Even though most microfibrils are longer than 500 μm , their small diameter (~ 20 μm) allows them to pass through the 200 and 500 μm mesh used in most neuston and manta nets [57].

According to us, the same reasoning can be made for the lines represented in the prediction model by class 4 (yellow dots), which, similar to the fibres, have been underestimated in the real situation due to the same problems of the fibres related to the sampling. In this case, due to their slightly larger diameter, the position they occupy in the two subplots (b) and (f) closely follows that of the fibres. The fragments could be represented by class 2 (blue dots), overlapping the other class in the subplot Figure 8b and very abundant and big in subplot Figure 8f and pellets by class 3 (green dots), present as outliers in both cases due to their numerical minority given by their origin [36,49,50].

The potential of our approach is its simple application and adaptability for a vast range of monitoring activities, such as the quantification and classification of macroplastics (>25 mm) in aquatic environments using aerial images. The algorithms easily adapt to

processing images from smartphones, handheld cameras, fixed observatories, and manned aerial and space platforms.

5. Conclusions

This paper presents a method to count and classify microplastic particles in dimensional and morphological classes, exhibiting promising results. The method makes use of both Computer Vision techniques and Machine Learning algorithms.

It is possible to automatically count MPs and divide them into groups with considerable time savings based mainly on the early ML algorithm application.

Supervised classification showed great potential, providing at the same time shortcomings especially in the correct classification of some specific types of MPs, such as pellets, which, however, could be explained by their specific origin, which makes them numerically lower than other categories of MPs.

The unsupervised classification provided useful indications of the sample's representativeness and probably a more realistic vision of the truth, suggesting that the MPs' smallest sizes are most likely missing and underestimated due to a lack in the sampling methodology. Discovering hidden features to understand the story unfolding in the picture is one of the goals of the unsupervised classification.

The possibility of using other machine learning techniques for the reconstruction of missing data, e.g., in the polymer characterisation field, will be assessed in the future. Our next goal is to improve and refine early ML algorithms to obtain predictions, using regression models, to better reconstruct the quantities of MPs over time, obviously based on a much more substantial dataset.

Author Contributions: Conceptualization, C.M. and C.C.; methodology, C.M. and C.C.; software, C.M.; validation, C.C.; formal analysis, C.M. and C.C.; investigation, C.C.; resources, C.M. and V.F.U.; data curation, C.C.; writing—original draft preparation, C.M. and C.C.; writing—review and editing, C.M. and C.C.; visualization, C.M. and C.C.; supervision, C.M. and V.F.U.; project administration, C.C. and V.F.U.; funding acquisition, C.M., C.C. and V.F.U. All authors have read and agreed to the published version of the manuscript.

Funding: This study was co-financed by INNOLABS Apulia Regional strategy for research and innovation guaranteed on Smart Specialization, a methodology that connects potential users with designers, for testing new products or services useful for solving specific problems of social relevance (Apulia POR FESR-FSE 2014-2020 Priority axis 1—Research, technological development, innovation).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors want to thank Nicola Palmisano for his precious support in the development of the Graphical User Interface (GUI).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Frias, J.P.G.L.; Nash, R. Microplastics: Finding a consensus on the definition. *Mar. Pollut. Bull.* **2019**, *138*, 145–147. [[CrossRef](#)]
2. Stock, F.; Kochleus, C.; Bansch-Baltruschat, B.; Brennholt, N.; Reifferscheid, G. Sampling techniques and preparation methods for microplastic analyses in the aquatic environment—A review. *TrAC Trends Anal. Chem.* **2019**, *113*, 84–92. [[CrossRef](#)]
3. Law, K.L.; Thompson, R.C. Microplastics in the seas. *Science* **2014**, *345*, 144–145. [[CrossRef](#)]
4. Thompson, R.C. Microplastics in the Marine Environment: Sources, Consequences and Solutions. In *Marine Anthropogenic Litter*; Bergmann, M., Gutow, L., Klages, M., Eds.; Springer International Publishing: Cham, Switzerland, 2015.
5. Klein, S.; Dimzon, I.K.; Eubeler, J.; Knepper, T.P. Analysis, Occurrence, and Degradation of Microplastics in the Aqueous Environment. In *Freshwater Microplastic*; Wagner, M., Lambert, S., Eds.; Springer International Publishing: Cham, Switzerland, 2018.
6. Campanale, C.; Massarelli, C.; Bagnuolo, G.; Savino, I.; Uricchio, V.F. The problem of microplastics and regulatory strategies in Italy. In *Plastics in the Aquatic Environment Stakeholders Role against Pollution*; Stock, F., Reifferscheid, G., Brennholt, N., Kostianaia, E., Eds.; Springer: Berlin/Heidelberg, Germany, 2019.

7. Li, J.; Liu, H.; Chen, J.P. Microplastics in freshwater systems: A review on occurrence, environmental effects, and methods for microplastics detection. *Water Res.* **2018**, *137*, 362–374. [[CrossRef](#)]
8. Campanale, C.; Stock, F.; Massarelli, C.; Kochleus, C.; Bagnuolo, G.; Reifferscheid, G.; Uricchio, V.F. Microplastics and their possible sources: The example of Ofanto river in southeast Italy. *Environ. Pollut.* **2020**, *258*. [[CrossRef](#)] [[PubMed](#)]
9. Chaczko, Z.; Wajs-Chaczko, P.; Tien, D.; Haidar, Y. Detection of Microplastics Using Machine Learning. In Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC), Kobe, Japan, 7–10 July 2019; pp. 1–8.
10. Gonçalves, G.; Andriolo, U.; Gonçalves, L.; Sobrai, P.; Bessa, F. Quantifying Marine Macro Litter Abundance on a Sandy Beach Using Unmanned Aerial Systems and Object-Oriented Machine Learning Methods. *Remote Sens.* **2020**, *12*, 2599. [[CrossRef](#)]
11. Lorenzo-Navarro, J.; Castrillón-Santana, M.; Gómez, M.; Herrera, A.; Marín-Reyes, P.A. Automatic counting and classification of microplastic particles. In Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods, Funchal, Portugal, 16–18 January 2018; pp. 646–652.
12. Lorenzo-Navarro, J.; Castrillon-Santana, M.; Santesarti, E.; De Marsico, M.; Martinez, I.; Raymond, E.; Gomez, M.; Herrera, A. SMACC: A System for Microplastics Automatic Counting and Classification. *IEEE Access* **2020**, *8*, 25249–25261. [[CrossRef](#)]
13. Wegmayr, V.; Sahin, A.; Sæmundsson, B.; Buhmann, J.M. Instance Segmentation for the Quantification of Microplastic Fiber Images. In Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass, CO, USA, 1–5 March 2020.
14. Pedrotti, M.L.; Petit, S.; Elineau, A.; Bruzaud, S.; Crebassa, J.-C.; Dumontet, B.; Martí, E.; Gorsky, G.; Cózar, A. Changes in the Floating Plastic Pollution of the Mediterranean Sea in Relation to the Distance to Land. *PLoS ONE* **2016**, *11*, e0161581. [[CrossRef](#)]
15. Schmidt, N.; Thibaul, D.; Galgani, F.; Paluselli, A.; Sempéré, R. Occurrence of microplastics in surface waters of the Gulf of Lion (NW Mediterranean Sea). *Prog. Oceanogr.* **2018**, *163*, 214–220. [[CrossRef](#)]
16. Gorsky, G.; Ohman, M.D.; Picheral, M.; Gasparini, S.; Stemman, L.; Romagnan, B.; Cawood, A.; Pesant, S.; Garcia-Comas, C.; Prejger, F. Digital zooplankton image analysis using the ZooScan integrated system. *J. Plankton Res.* **2010**, *32*, 285–303. [[CrossRef](#)]
17. Abramoff, M.; Magalhães, P.; Ram, S.J. Image Processing with ImageJ. *Biophotonics Int.* **2003**, *11*, 36–42.
18. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671–675. [[CrossRef](#)] [[PubMed](#)]
19. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, in press. [[CrossRef](#)] [[PubMed](#)]
20. Kedzierski, M.; Falcou-Préfol, M.; Kerros, M.E.; Henry, M.; Pedrotti, M.L.; Bruzaud, S. A machine learning algorithm for high throughput identification of FTIR spectra: Application on microplastics collected in the Mediterranean Sea. *Chemosphere* **2019**, *234*, 242–251. [[CrossRef](#)]
21. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
22. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *39*, 1137–1149. [[CrossRef](#)]
23. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. *arXiv* **2014**, arXiv:1409.4842.
24. Beyeler, M. *Machine Learning for OpenCV*; Packt Publishing Ltd.: Birmingham, UK, 2017.
25. Sebe, N.; Cohen, I.; Garg, A.; Huang, T.S. *Machine Learning in Computer Vision*; Springer: Dordrecht, The Netherlands; Berlin/Heidelberg, Germany; New York, NY, USA, 2005; ISBN 1402032749.
26. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **1958**, *65*, 386–407. [[CrossRef](#)]
27. Michalski, R.S.; Carbonell, J.G.; Mitchell, T.M. (Eds.) *Machine Learning: An Artificial Intelligence Approach*; Morgan Kaufmann: Los Altos, CA, USA, 1986.
28. Rowley, H.; Baluja, S.; Kanade, T. Neural network-based face detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 23–38. [[CrossRef](#)]
29. Pulli, K.; Baksheev, A.; Korniyakov, K.; Eruhimov, V. Realtime Computer Vision with OpenCV. *Queue* **2012**, *10*, 40–56. [[CrossRef](#)]
30. Rosebrock, A. *Practical Python and OpenCV. An introductory, Example Driven Guide to Image Processing and Computer Vision*, 3rd ed. ASIN B078QFL1NY. 2017. Available online: PyImageSearch.com (accessed on 15 June 2021).
31. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [[CrossRef](#)]
32. Sankur, B. Survey over image thresholding techniques and quantitative performance evaluation. *J. Electron. Imaging* **2004**, *13*, 146–168. [[CrossRef](#)]
33. Hu, M.K. Visual Pattern Recognition by Moment Invariants. *IRE Trans. Info. Theory* **1962**, *8*, 179–187.
34. Green, G. An essay on the application of mathematical analysis to the theories of electricity and magnetism. *arXiv* **1828**, arXiv:0807.0088.
35. Jurkat, W.B.; Nonnenmacher, D.J.F. The General Form of Green’s Theorem. *Proc. Am. Math. Soc.* **1990**, *109*. [[CrossRef](#)]
36. Helm, P.A. Improving microplastics source apportionment: A role for microplastic morphology and taxonomy? *Anal. Methods* **2017**, *9*, 1328–1331. [[CrossRef](#)]
37. Lusher, A.L.; Welden, N.; Sobral, P.; Cole, M. Sampling, isolating and identifying microplastics ingested by fish and invertebrates. *Anal. Methods* **2017**, *9*, 1346–1360. [[CrossRef](#)]

38. Dyachenko, A.; Mitchell, J.; Arsem, N. Extraction and identification of microplastic particles from secondary wastewater treatment plant (WWTP) effluent. *Anal. Methods* **2017**, *9*, 1412–1418. [[CrossRef](#)]
39. Hidalgo-Ruz, V.; Gutow, L.; Thompson, R.C.; Thiel, M. Microplastics in the Marine Environment: A Review of the Methods Used for Identification and Quantification. *Environ. Sci. Technol.* **2012**, *46*, 3060–3075. [[CrossRef](#)] [[PubMed](#)]
40. Aha, D.W.; Kibler, D.; Albert, M.K. Instance-based learning algorithms. *Mach. Learn.* **1991**, *6*, 37–66. [[CrossRef](#)]
41. Naomi, S.A. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **1992**, *46*, 175–185. [[CrossRef](#)]
42. Belur, V. *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*; IEEE Computer Society Press: Washington, DC, USA, 1991.
43. Shakhnarovich, G.; Darrell, T.; Indyk, P. (Eds.) *Nearest-Neighbour Methods in Learning and Vision*; The MIT Press: Cambridge, MA, USA, 2005; ISBN 026219547X.
44. Pedregosa, F.; Voroquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
45. Davis, J.; Goadrich, M. The relationship between Precision-Recall and ROC curves. In Proceedings of the 23rd International Conference on Machine Learning (ICML), Pittsburgh, PA, USA, 25–29 June 2006; pp. 233–240.
46. Flach, P.A.; Kull, M. Precision-Recall-Gain Curves: PR Analysis Done Right. In Proceedings of the 29th Conference on Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 7–12 December 2015.
47. GESAMP. *Report and Studies No. 99. Guidelines for the Monitoring and Assessment of Plastic Litter in the Ocean*; United Nations Environment Programme (UNEP): Nairobi, Kenya, 2019; ISSN 10204873.
48. Quinn, G.P.; Keough, M.J. *Experimental Design and Data Analysis for Biologists*; Cambridge University Press: Cambridge, UK, 2002.
49. Turner, A.; Wallerstein, C.; Arnold, R. Identification, origin and characteristics of bio-bead microplastics from beaches in western Europe. *Sci. Total Environ.* **2019**, *664*, 938–947. [[CrossRef](#)] [[PubMed](#)]
50. Boucher, J.; Friot, D. *Primary Microplastics in the Oceans: A Global Evaluation of Sources*; IUCN: Gland, Switzerland, 2017; p. 43.
51. Zhu, Y.; Yeung, C.H.; Lam, E.Y. Digital holographic imaging and classification of microplastics using deep transfer learning. *Appl. Opt.* **2021**, *60*, 38–47. [[CrossRef](#)]
52. Peršak, T.; Viltušnik, B.; Hernavs, J.; Klančnik, S. Vision-Based Sorting Systems for Transparent Plastic Granulate. *Appl. Sci.* **2020**, *10*, 4269. [[CrossRef](#)]
53. Hufnagl, B.; Steiner, D.; Renner, E.; Löder, M.G.J.; Laforsch, C.; Lohninger, H. A methodology for the fast identification and monitoring of microplastics in environmental samples using random decision forest classifiers. *Anal. Methods* **2019**, *11*, 2277–2285. [[CrossRef](#)]
54. Bagaev, A.; Mizyuk, A.; Khatmullina, L.; Isachenko, I.; Chubarenko, I. Anthropogenic fibres in the Baltic Sea water column: Field data, laboratory and numerical testing of their motion. *Sci. Total Environ.* **2017**, *599*, 560–571. [[CrossRef](#)]
55. Barrows, A.P.W.; Cathey, S.E.; Petersen, C.W. Marine environment microfiber contamination: Global patterns and the diversity of microparticle origins. *Environ. Pollut.* **2018**, *237*, 275–284. [[CrossRef](#)]
56. Suaria, G.; Achtypi, A.; Perold, V.; Lee, J.R.; Pierucci, A.; Bornman, T.G.; Aliani, S.; Ryan, P.G. Microfibers in oceanic surface waters: A global characterization. *Sci. Adv.* **2020**, *6*. [[CrossRef](#)] [[PubMed](#)]
57. Ryan, P.G.; Suaria, G.; Perold, V.; Pierucci, A.; Bornman, T.G.; Aliani, S. Sampling microfibres at the sea surface: The effects of mesh size, sample volume and water depth. *Environ. Pollut.* **2019**, *258*. [[CrossRef](#)]