IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 3 June 2022, accepted 30 June 2022, date of publication 7 July 2022, date of current version 18 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3189012

RESEARCH ARTICLE

Automatic Pollen Classification and Segmentation Using U-Nets and Synthetic Data

MIHAI BOLDEANU[1], MÓNICA GONZÁLEZ-ALONSO[2], (Student Member, IEEE), HORIA CUCU[1], (Member, IEEE), CORNELIU BURILEANU[1], (Life Senior Member, IEEE), JOSE MARIA MAYA-MANZANO[3], AND JEROEN TITUS MARIA BUTERS[3]

[1]Speech and Dialogue Research Laboratory, Politehnica University of Bucharest, 060042 Bucharest, Romania
[2]Environmental Biology Department, University of Navarra, 31009 Pamplona, Spain
[3]Center of Allergy and Environment (ZAUM) and Helmholtz Center, Technical University of Munich, 80333 Munich, Germany

Corresponding author: Mihai Boldeanu (mishuboldeanu@gmail.com)

This work was supported in part by the Grant of the Romanian Ministry of Education and Research, CCCDI—UEFISCDI, under Project PN-III-P2-2.1-PED-2019-2278, within PNCDI III.

ABSTRACT Pollen allergies have become one of the most wide-spread afflictions that impact quality of life. This has made automatic pollen detection, classification and monitoring a very important topic of research. This paper introduces a new public annotated image data-set of pollen with almost 45 thousand samples obtained from an automatic instrument. In this work we apply some of the best performing convolutional neural networks architectures on the task of pollen classification as well as some fully convolutional networks optimized for image segmentation on complex microscope images. We obtain an F1 scores of 0.95 on the new data-set when the best trained model is used as a fully convolutional classifier and a class mean Intersection over Union (IoU) of 0.88 when used as an object detector.

INDEX TERMS BAA-500, pollen classification, pollen image segmentation, U-net, artificial pollen dataset.

I. INTRODUCTION

Pollen allergies cause most of the seasonal allergies [1]. Because the number of people affected by these sort of ailments is on the rise there is a lot of interest in developing methods of predicting atmospheric pollen concentrations for plants known to cause issues.

There are two ways in which pollen monitoring can improve the quality of life for most allergic people. Nowcasting can be used to send alerts, similar to how extreme weather alerts are broadcast. This information could prove useful in minimizing exposure to pollen, from certain species of plants. Another, more complex solution, would be creating pollen forecasts similar to how weather prediction functions. Both solutions can work only if reliable historical data is available. These data sets have to be gathered using standardized equipment and methodologies such as automatic pollen monitoring systems.

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

To create predictive models for pollen from multiple species there is a need for accurate pollen counts over multiple seasons. This is needed to find a significant statistical relationship between pollen and other meteorological parameters such as temperature, humidity, pressure, or other environmental variables. All of these sources of information can then be used in developing time-series analysis that can predict days or even weeks in advance the presence of pollen, from a certain species, at a certain location.

The issue in developing forecast models is that pollen counting so far has been mainly a human-centric activity, with a time-intensive procedure where technicians or Ph.D. students are used to count and classify the particles captured by Hirst traps [2] or similar devices [3]. These methods of getting pollen information have some inherent limitations. The main issue that comes to mind is the lack of scalability of systems that rely on human operators. Another aspect is the lag between when pollen is captured by the device and when the information actually reaches the general public, with most pollen counting happening at the end of the week for the preceding week. Because different groups use

different methodologies for counting and classifying pollen there is an inherent bias in the pollen data from different locations.

Many papers have tackled the task of pollen classification in the past, unfortunately, most of them have the same weakness, they rely on hand-crafted data-sets to train machine learning models. This limits the use of the trained models to data sources that use the exact same pollen sample preparation steps as the the hand-crafted data sets. This makes these models less useful in real-world applications, where there are variations in the pollen sample processing.

To alleviate the issues associated with manual pollen monitoring, automated systems have become more widespread. One fully automatic system is the BAA-500 device [4]. This instrument operates similar to the classical Hirst trap but eliminates the human input in preparing the microscope slides and selecting focus levels. This makes measurements much more standardized and part of the same data distribution, making them inter-compatible for use in modeling. While such devices are great for obtaining raw data, there is still a need for more algorithms development to fully exploit these data sources and to go from images to sufficiently reliable pollen counts that would be then useful for modeling or for creating alerts.

In normal operation, a BAA-500 device will generate images of size 1280×960 that contain a large number of different particles. Before any classification can occur, the image must be segmented at a particle by particle level. While the instrument manufacturers provide some algorithms to do this, the complex shape of the particles makes it very difficult to accurately segment such images using classical image processing methods.

In this work, we tackle this problem by utilizing a fully convolutional neural network architecture to directly segment the large images and make classifications in one step. This model is trained using synthetic images created from a new public data set of pollen.

The paper is structured as follows: Section 2 presents related work to the use of convolutional neural networks for pollen classification, and more specifically looks at results obtained on data from the BAA-500 device. Section 3 presents an overview of the new annotated pollen data-set from a BAA-500 device and some useful augmentation strategies; also in this section the proposed architecture for pollen segmentation is presented. In the results section, we compare multiple approaches for pollen classification. Finally, the conclusions present the advantages and the limitations of our proposed system.

Our Contributions:

The main contributions presented in this work are:

- A new public pollen data-set obtained from a BAA-500 device that was first processed by automated means and then manually validated by human experts. This data-set contains over 44000 cropped images of 19 classes (16 pollen types, 2 fungus spore types and a junk class) of particles.
- Evaluation of a multitude of CNNs architectures on the data-set to obtain a classification score baseline.
- Development of a method of training fully convolutional models to be able to segment and classify pollen from automated BAA-500 system.

II. RELATED WORK

The need for automatic pollen classification was first identified in the late '60s by [5], with the first attempts done by [6] at the end of the '70s with the use of electron microscopy to obtain high-detail images of pollen and use of discriminant analysis as the tool for classification. The use of an electron microscope to obtain pollen images used for classification was completely impractical due to the requirements of obtaining such images and was only used as a proof-of-concept. After this, the move was towards visible light microscopy to obtain pollen images used for classification usually with texture analysis methodologies, such as those done by [7], [8] and [9].

Starting with the 2000s, as the field of machine learning had expanded and matured, new techniques have been used for pollen classification with many attempts such as [10]–[12] and [13].

A. POLLEN DATA-SETS

To further the work done on pollen classification a number of pollen data-sets, containing annotated images of pollen samples, appeared and among them some notable examples are:

- POLEN23E that first appeared in [14] is a small set containing 805 pollen images from 23 classes. The authors that introduced this data-set obtained a Correct Classification Rate (CCR) of 64%.
- POLLEN73S introduced in [15] is a medium sized set containing 2523 images from 73 classes. In [15] the authors achieved a precision results of 95.7% and 94.0% when using DenseNet-201 and ResNet-50 models.
- Finally, POLLEN13K was presented in [16] and is a larger set containing over 13000 images but from only 5 classes.

While each of these data-sets is very useful in developing new and better approaches for automatic pollen classification, they all suffer from the same limitation. This being that the data-sets are hand-crafted, with a lot of human intervention, and not built in particular standardized way. All the above data-sets include pollen preparation steps that have to be manually done by technicians, like selecting a pigment to highlight different aspects of the pollen particle and selecting different focus levels for each pollen type to have the sharpest image. This introduces a lot of bias that makes models trained on such data-sets unusable on new samples from automated pollen capture systems. This makes their applicability limited when increasing the scale to automated systems. To use a model trained on such data-sets the data pipeline would have to include all of the pre-processing steps done by humans,

which would defeat the purpose of having an automated system to gather the new samples.

In recent years, with the advancement of automatic particle analyzers, some data-sets have been created using only data from completely automated systems. Some examples of such data-sets are:

- Data-set-15 [17] contains overall 51,277 samples and 15 classes and the originators obtained an unweighted average precision of 83.0 % and an unweighted average recall of 77.1 % across 15 classes of pollen taxa.
- Data-set-31 [18] is an expanded version of Data-set-15 but with some extra classes. In [18] the authors achieved an unweighted average F1 score of 93.8% across 15 classes and an unweighted average F1 score of 75.9% across 31 classes. While the result on these data-sets are very promising, the main issue still remains that these classifiers rely on classical image processing methods for segmentation. And these segmentation methods are not optimal for complex patterns formed by pollen particles on microscope slides.

While Data-set 15 and 31 are great resources they are not publicly available and this limits the amount of research that can be done using them. Our Data-sets addresses most of the issues of previously published sets because it is publicly available and obtained from an automated system.

B. POLLEN CLASSIFICATION ALGORITHMS

Some of the classical machine learning approaches previously used for pollen classification are: Support Vector Classifiers (SVCs) used in [19], Random Forest (RF) in [20], Decision Trees with Adaboost ensemble used in [21], and Multi-Layer Perceptron (MLP) in [13]. Because such methods have difficulties when applied to images, dimensionality reduction has to be first applied. Some dimensionality reduction used with pollen images are the histogram of gradients (HOG) [22] and local binary pattern (LBP) [23] as previously used for pollen classification in [16].

Convolutional neural networks have become, over the last decade, the solution when dealing with computer vision tasks. This is true also for pollen classification with most attempts using such networks. Because training such large models requires a huge amount of data that is not always easy to obtain, the standard approach is to use a pre-trained model to jump-start the training on a new vision task.

Because we introduce a new pollen data-set in this paper, we run the ''standard'' battery of deep learning architectures specialized on computer vision, including VGG-16 [24], VGG-19 [24], ResNet50 [25], Inception V3 [26], Xception [27] and DenseNet201 [28]. All these models were selected because they showed very good results in recent pollen classification work [15] and [18].

III. MATERIALS AND METHODS

In this section the pollen data-set is presented along with all of the pre-processing and augmentation methods used.

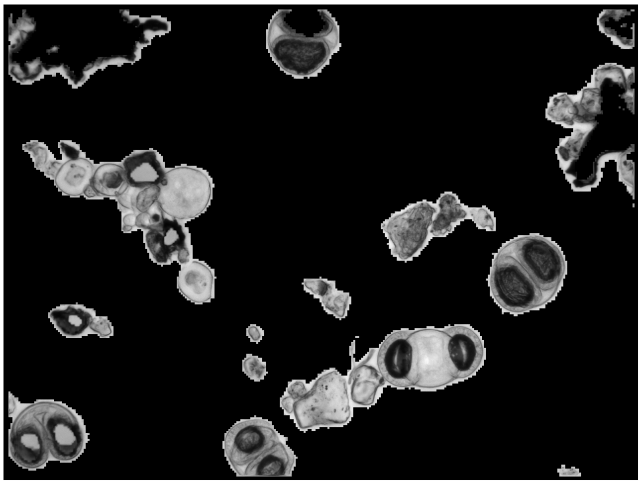


FIGURE 1. The full image of pollen with dirt and other contaminants from BAA-500 device. Rough segmentation made by instrument.

The architecture selection and the experimental setup is also discussed in this section.

A. DATA-SET OVERVIEW

The BAA-500 instrument provides microscope images of pollen at multiple focus levels. These images are merged by the device into a synthetic image to minimize storage requirements. These synthetic images are then roughly segmented by the device as can be seen in Figure 1, with mostly empty regions being replaced by a black background.

The device manufacturers also provide a framework for individual particle segmentation and classification using several hand-crafted features. Using this as a starting point, a database of pollen, representative for Europe, was compiled from many devices. A sub-set of this database, validated by human experts, is the basis of the data-set introduced in this work.

In Table 1 the number of samples from each pollen type is presented. While most of the classes are representative of the different genera of plants, some such as *Alternaria*, Fungus, and Junk are other objects that can be found on the microscope slides alongside pollen.

The data-set is relatively balanced compared to previously published data-sets (Data-Set-15–31). Still some class imbalance is present and that has to be accounted for when training a classification model.

This new data-set is completely ''compatible'' with the other sets obtained from a BAA-500 device and a larger mixed data-set could be considered by merging it with the previously discussed ones.

Moreover, thanks to the lack of human involvement in the preparation of the slides and in taking the pollen images, there is a good opportunity to use a model trained on this data-set on real-world data from BAA-500 devices, such as those in the Bavaria monitoring network. While this data-set was obtained completely automatically, some biases were still

TABLE 1. Data-set overview.

GENUS	NUMBER OF SAMPLES	% OF DATA-SET
ALNUS	2860	6.38 %
ALTERNARIA	1849	4.11%
AMBROSIA	143	0.31%
ARTEMISIA	791	1.76%
BETULA	2894	6.44%
CARPINUS	2871	6.39%
CORYLUS	220	0.48%
CUPRESSACEAE	2702	6.01%
FAGUS	2958	6.58%
FRAXINUS	2934	6.53%
FUNGUS	3458	7.70%
JUNK*	5297	11.59%
PINUS	1830	4.07%
PLANTAGO	3034	6.75%
PLATANUS	828	1.84%
POACEAE	2390	5.32%
POPULUS	2958	6.58%
QUERCUS	2806	6.24%
URTICACEAE	2085	4.64%

* not pollen or fungus genus, but representative for dust or other particles that might appear on microscope slides.

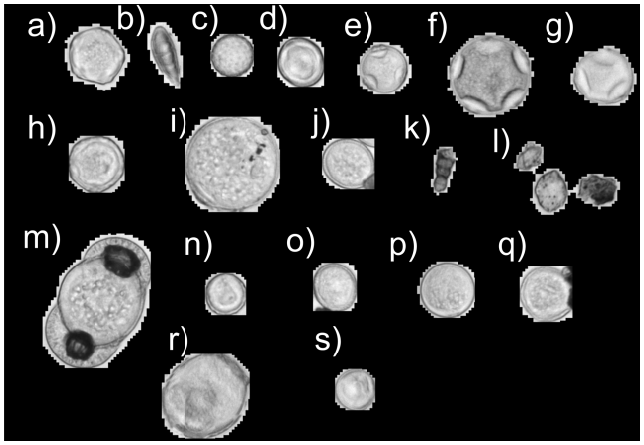


FIGURE 2. All the particle types present in the database: a) Alnus, b) Alternaria, c) Ambrosia, d) Artemisia, e) Betula, f) Carpinus, g) Corylus, h) Cupressaceae, i) Fagus, j) Fraxinus, k) Fungus, l) Junk, m) Pinus, n) Plantago, o) Platanus, p) Poaceae, q) Populus, r) Quercus, s) Urticaceae.

imparted by the human experts that made the re-classification and validation of the data.

B. DATA ANALYSIS AND AUGMENTATION

The data-set is composed of individually cropped images of pollen and other particles with a maximum size of 360 × 360 pixels and grayscale color information. Examples of the types of particles present in the data set are shown in Figure 2. Due to the different sizes of the particles, padding was used to bring all of them to the same shape. This was easily done because the background was already black.

From Table 1 we can see that there is some class imbalance in the data-set that has to be accounted for. This can be to a degree resolved by over-sampling the less predominant classes and by applying augmentations to them.

To the original images, several types of transformations were applied to increase the number of samples available for training and, to a lesser degree, prevent over-fitting.

Positional augmentations such as vertical and horizontal flipping, rotating, and shearing of the images was first applied. The next step involved applying color augmentations such as modifying the brightness, contrast, or histogram shifting. And lastly, we applied gray patches overlaid on the images to occlude parts of the original sample. This last step was done to force the model to rely on the entire image for context and not just particular features present in the pollen.

All of the augmentations were applied either when training a classifier or when constructing the samples used to train the U-net models used for class segmentation. As such, the images generated through augmentation are not part of the data-set, available on request, but the source code used for augmentation is provided at the end of the paper.

C. PROPOSED ARCHITECTURE

The machine learning task approached in this work is that of image segmentation and classification. For such a task, U-nets [29] are a very common network architecture used. This is in part thanks to the ability of such networks to learn complex patterns and be able to segment organic patterns present in images.

Since they were initially proposed [29], U-nets have become a staple in many scientific fields in the task of class segmentation of complex tissues [30], cellular components [31], satellite image processing [32] and even geological studies [33].

While many improvements have been applied to U-net, the basic concept of a fully convolutional neural network is still powerful on its own and the U-net is still the go-to architecture when approaching a new segmentation task.

1) U-NET AND VARIANTS

The U-net architectures used in this work are based on the networks first presented in [29]. Starting with the original U-net a few variants, based on work done by [34] and [35], were also tested to find the best performance on our data-set. Considering that U-nets are, at their core, nothing more than a Fully Convolutional Network (FCN) [36] with long-distance skip connections, we tested several different configurations to find the best-performing model for our task.

As presented in Figure 3 we started with the FCN pictured in a) and U-net in b). After this we replaced concatenation with addition, for the long-distance skip connection in the version shown in c). So instead of concatenating the information, from the contraction path to the expansive side, we just added them together at the appropriate levels. Another modification applied was adding more skip-connections that skip only one convolution block at a time d). The last two variants are the combinations of b) and c) with the d) variant.

For all the U-net networks variants the depth was kept fixed using only 3 max-pooling layers [37] and 3 up-sample layers. The number of convolution filters was always doubled

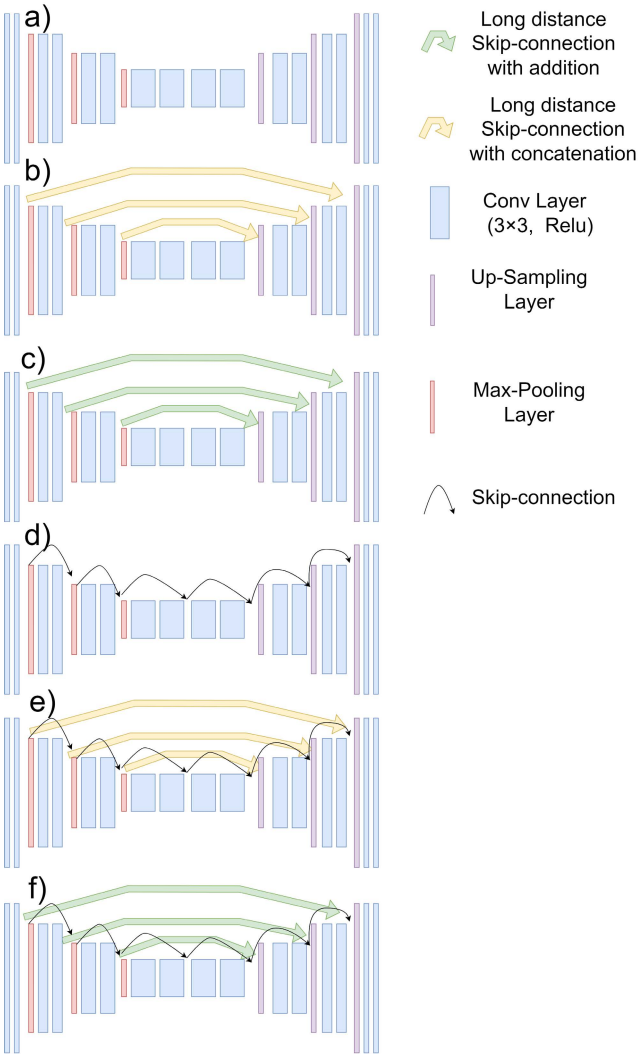


FIGURE 3. The fully convolutional networks used in this paper: a) FCN, b) U-net, c) U-net with addition instead of concatenation, d) FCN with skip connections, e) U-net with skip connections, f) U-net with addition instead of concatenation and with skip connections.

after a max-pooling operation. This means that the number of kernels of the first layer defines the width of the entire network. This was varied with 4, 8, 16, and 32 selected as the number of kernels in the first layer.

The activation function selected was that of Rectified linear units or ReLu [38] and Batch Normalization layers [39] were used to aid in training. The order of layers used at each level is the convolution layer, followed by the batch normalization layer, followed by the non-linear activation.

2) BUILDING TRAINING DATA FOR U-NET

In order to train a fully convolutional neural network, labels for images are not sufficient. The training labels have to be extended at the pixel level for each sample in the data-set. Because our goal is to train a model capable of segmenting images into masks for each type of pollen, the training data has to be adequately prepared.

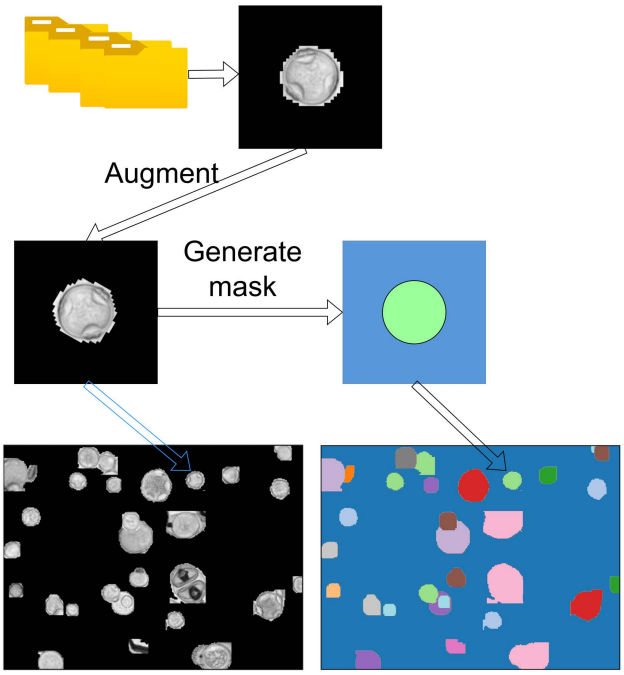


FIGURE 4. Cropped images are randomly selected from the data-set, augmented and used to create large synthetic images similar in complexity to the real images produced by the BAA-500.

Using the database presented above, we created synthetic images similar in size and complexity to the original large images produced by the BAA-500. This is useful because the fully trained model can then be used directly on the output of the pollen monitoring instrument. The model will either be used to identify the presence of certain particles and count them. Or even in an easier use case, the model could be used to create masks that can be used to remove unwanted classes from an image (e.g. junk class).

All of the augmentation techniques, presented in the previous section, were used on the small cropped images that were then used as building blocks, for creating the large synthetic images. Because the junk class is much more varied compared to the pollen classes, many junk backgrounds were used to train the models to classify as junk vegetable fibers, commonly found in alongside pollen, that were not present in the data-set due to their large size. The junk backgrounds were treated as normal backgrounds and particles were overlaid to create a complex image.

When generating each synthetic image, we started with either a fully black background or a junk background. After this, we randomly selected a class of particles and selected a random sample from that class. This was then randomly augmented with one or more transformations and overlaid with the background image. At the same time, the ground-truth mask was constructed as presented in Figure 4.

The training sub-set was used to create random large synthetic images during training and a seed based creation of large synthetic images from the validation set was used to

TABLE 2. Classification F1 score for classical ML approaches.

Model type/Feat.Eng.	HOG	LBP
Linear SVC	0.46	0.46
RBF SVC	0.29	0.48
Random Forest	0.46	0.53
ADABoost	0.41	0.52
MLP	0.47	0.62

TABLE 3. Classification F1 score for deep architectures.

Architecture used	Pre-trained on ImageNet	Fully train on pollen
VGG-16	0.82	0.90
VGG-19	0.80	0.92
ResNet50	0.59	0.90
InceptionV3	0.85	0.93
Xception	0.86	0.93
DenseNet201	0.87	0.92

evaluate from epoch to epoch. This was done to maintain a fixed target for the algorithm to learn towards.

IV. RESULTS AND DISCUSSIONS

This section presents all of the experiments done to find a good classifier system for the data from BAA-500 devices. The objective metric used to compare classification models is an unweighted F1 score and for the segmentation task, the metric used was the unweighted mean IoU for all of the classes.

Before training any model, the data-set was split into 3 parts, the training set containing 80%, the validation set with 10%, and the test set with 10% of the original data-set. The split was done proportional to the number of samples per class in most cases. For the classes that had very few examples a minimum of 20 were reserved for testing.

A. CLASSICAL APPROACHES

For classical approaches to pollen classification we started from the work done by [16].

The data was first re-scaled from the 0–255 range to the 0–1 range to make it easier for the models to train. Our cropped samples were 360×360 in size and a dimensionality reduction was necessary to accomplish this. We used histogram of gradients (HOG) and local binary pattern (LBP) as two different feature engineering steps.

Because they showed good results in previous works [16] some architectures were selected including Support Vector Classifiers (SVCs), Random Forest (RF), Decision Trees with Adaboost ensemble, and Multi-Layer Perceptron (MLP). All of the selected models come from the sci-kit learn packages and were used with default settings.

As presented in Table 2, the performance of classical methods is limited. This could be explained by some things such as a loss of information when applying dimensionality reduction methods(HOG and LBP), the simplicity of the models used and the complexity of the task. This was expected, but it is always best to try the simplest solution to a problem first.

B. DEEP ARCHITECTURES

The experiments with CNN architectures are split into two groups: i) using a pre-trained model and fine-tuning or ii) training from scratch. When using the pre-trained model the gray-scale images are treated as RGB images with identical information on all of the channels, to have the same input shape as the models trained on Imagenet [40]. The fully connected layers, on top of the convolutional part, of each model are replaced with a Global Pooling layer and a final soft-max layer used for classification. The convolutional layers are frozen and only the small fully connected part of the network is trained. The results are good and quite close for most of the models with an exception being the ResNet50 that suffers the most when pre-trained.

When moving to train the models from scratch, the input of the networks is considered as a $360 \times 360 \times 1$ tensor because there is no point in duplicating the grayscale image to three channels. Another difference is that all the layers are randomly initialized and all of them are trainable.

All of the networks in these experiments were trained for 100 epochs with a batch size of 32. The augmentations were randomly applied to images during training. The results presented were on the test set and for the best version of each model obtained by saving the model weight after each epoch, where performance increased on the validation set.

We obtained good and quite close results for all of the model architectures, with the worst performance being for the VGG-16 and ResNet50. This might be in part to the fact that we used a Global Pooling layer [41] instead of multiple dense layers after the convolutional part of the network. The very poor performance of the pre-trained ResNet50 might indicate that the model focuses a lot on color when making classifications.

While an in-depth hyper parameter search could improve the accuracy of these CNNs, the main limitation of relying on other methods for segmentation would still be present. This brings us to the final set of experiments presented in this work.

C. U-NET AND VARIANTS

Moving forward from the baseline of the CNN architectures used, we get to the U-net experiments.

These runs were done for all 6 variants of the U-net presented in the Materials and Methods section. For each variant of the U-net, we ran 4 different experiments with the number of filters of the first layer changed. This was done to find the optimum in the number of convolution kernels to have good performance of the models while not going into an overfitting regime.

Looking over the graphs in Figure 5 that show the training over 100 epochs for all the models at a certain width, we see that for the models with widths 4 and 8 have a validation score higher than the training score, meaning that it is easier for the

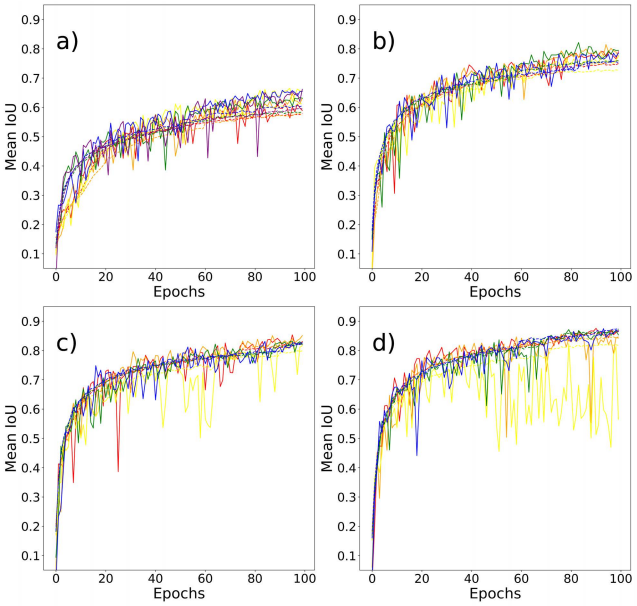


FIGURE 5. The training for all FCNs (all 6 variants): a) width 4, b) width 8, c) width 16, d) width 32. The segmented line is for train score and the continuous line is for validation score.

TABLE 4. Classification class mean unweighted class IoU for U-net and variants at different widths.

U-NET VARIANT	W-4	W-8	W-16	W-32
FCN	0.61	0.72	0.71	0.79
U-NET	0.67	0.81	0.85	0.86
U-NET(ADD)	0.65	0.79	0.85	0.88
FCN + RES	0.64	0.78	0.82	0.82
U-NET + RES	0.66	0.78	0.83	0.87
U-NET(ADD) + RES	0.66	0.82	0.83	0.88

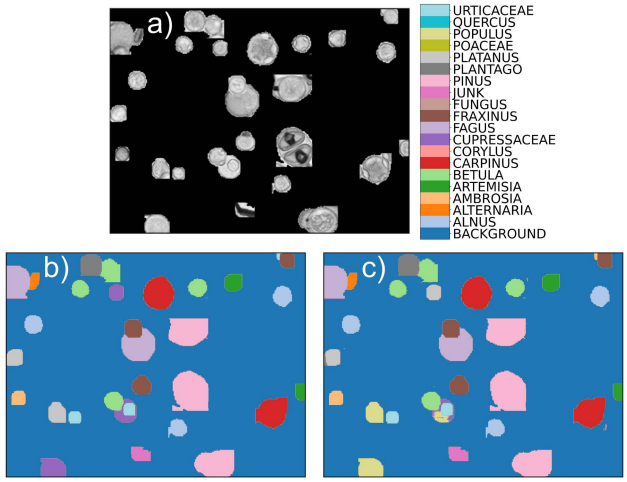


FIGURE 6. The results of the best performing U-net on a synthetic image: a) The synthetic image, b) the ground truth, c) the model prediction.

model to classify the data-set when it is not augmented than when it is augmented. At a width of 16 we see that the training and validation score is very close meaning that the model is well parameterized. And finally, when going at the largest

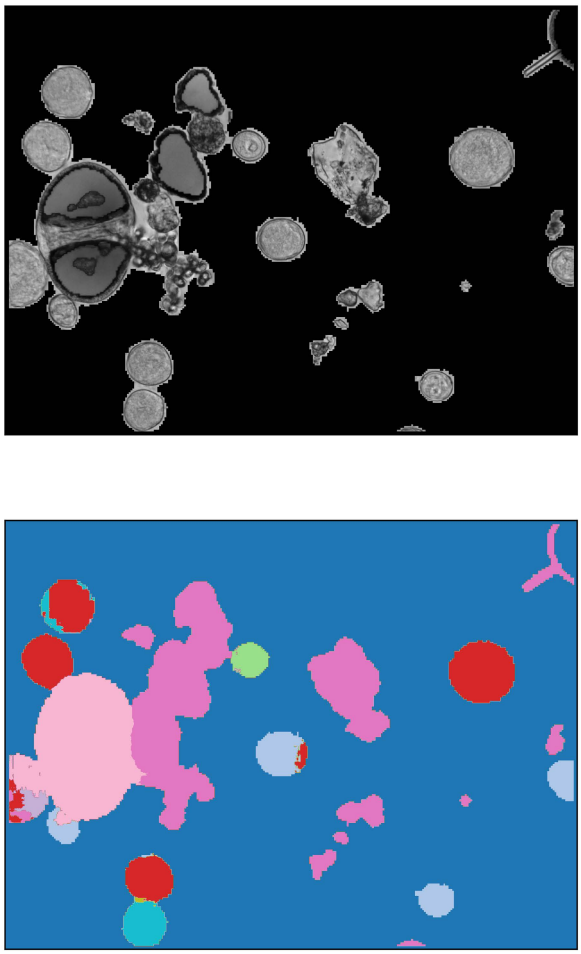


FIGURE 7. The results of the model when used on real data from BAA-500 device. Top image is the original gray-scale image and bottom one is the prediction. The model manages to differentiate quite well between different types of pollen and also manages to correctly identify the junk in the image.

width of 32 we see that some over-fitting is present and we have diminishing returns.

Because we could generate the images used in training the U-net, programmatically we used 1000 images per training epoch and generated new ones after each epoch. The validation images, used to evaluate the model after each epoch, were generated controlling the random seeds to have the same images at each evaluation.

To have a comparison between the U-net model and the other CNN architectures used for classification we propose an aggregation method to reduce an output mask to just one label. The output mask of the U-net is added on the x and y spatial dimensions leaving a $1 \times 1 \times 20$ vector with the highest value in the class that is most present in the image. The first element of the vector is discarded because it is used for the background.

Using this approach we obtained a classification average unweighted F1 score of 0.95 for the U-net. This is a good way of validating that the model learned more from the pollen particles than other approaches.

TABLE 5. The time required to train and the number of trainable parameters for each model used.

Model	Time to train	Number of parameters
Linear SVC (HOG)	474 seconds	12,960
RBF SVC (HOG)	380 seconds	110,979
Random Forest (HOG)	16 seconds	N/A
ADABoost (HOG)	302 seconds	N/A
MLP (HOG)	91 seconds	66,819
Linear SVC (LBP)	65 seconds	600
RBF SVC (LBP)	23 seconds	5,130
Random Forest (LBP)	7 seconds	N/A
ADABoost (LBP)	190 seconds	N/A
MLP (LBP)	35 seconds	4,919
VGG-16 (Pre-trained)	4 min/epoch	9,747
VGG-19 (Pre-trained)	4.5 min/epoch	9,747
ResNet50 (Pre-trained)	4 min/epoch	38,931
InceptionV3 (Pre-trained)	4 min/epoch	38,931
Xception (Pre-trained)	4 min/epoch	38,931
DenseNet201 (Pre-trained)	5 min/epoch	36,499
VGG-16	11 min/epoch	14,723,283
VGG-19	13 min/epoch	20,032,979
ResNet50	10 min/epoch	23,567,251
InceptionV3	7 min/epoch	21,806,707
Xception	17 min/epoch	20,845,307
DenseNet201	15 min/epoch	18,123,155
FCN (32 width)	7 min/epoch	2,940,596
U-net (32 width)	7 min/epoch	3,370,676
U-net(add) (32 width)	7 min/epoch	2,983,604
FCN + Res (32 width)	7 min/epoch	3,092,148
U-net + Res (32 width)	7 min/epoch	3,565,236
U-net(add) + Res (32 width)	7 min/epoch	3,135,156

Finally, this architecture has the advantage of working on the full images from the device. In Figure 6 we can observe the segmentation results on a synthetic image. In the first part, we see that while the particle identification is very good the main problem comes when classifying some pollen types from under representative classes.

When applying the network to real data, as in Figure 7, we get good results although there is some difficulty in finding clear boundaries between particles. This is caused by the fact that the cropped images, even when overlapped in our synthetic data, had clear boundary artifacts that the model learned to take advantage of.

The results on real images are promising considering the use of a U-net for so many classes. The possible use cases for this model could be either the full use as an object detector or an easier use case of just creating a mask for junk and eliminating such particles from the images to have an easier time at using other segmentation approaches.

A comparison of the time complexity and the memory complexity of all of the models presented in this section is highlighted in Table 5. The classical machine learning models can be trained very fast when compared to the deep learning architectures. This is not a real advantage as training happens only once and the performance is significantly worse. The U-net architecture is about six times smaller, in terms of the number of parameters, but performs equal or better than the best reference models, in the classification task. This might be because training for a more difficult task, segmentation, makes better use of the training data-set than training directly for classification. The training is longer for the U-net architectures because the training samples are created at training time.

V. CONCLUSION

This work focuses on three topics regarding pollen classification.

First, we introduce a new large public pollen data-set. This data-set was obtained from an automated particle monitoring system and is useful in developing models that can run operationally, because there are no pollen preparation steps that need human intervention.

Second, we evaluate the "standard battery" of classical machine learning and convolutional neural networks for image classification on our data-set and obtain good results that can be viewed as a baseline for future work.

Finally, we propose a method for training a fully convolutional network on synthetic data created from our data-set. When this model is used as a classifier it gets unweighted mean F1 score of 0.95, surpassing all of the standard convolutional networks. The model trained on the synthetic data is then used to segment images, from the automated device, and the results are promising. This is particularly important because the convolutional network proposed solves both segmentation and classification in one pass, as opposed to baseline algorithms which only address classification and need to be supplemented with a different segmenting pipeline.

The next steps in validating our results is to do an intercomparison campaign with both Hirst type devices, still considered the standard in pollen counting, and a BAA-500 device at the same location for an entire pollen season. This step is crucial in convincing the pollen counting community of the usefulness of these new approaches.

The code to build, train and deploy the models and the code used to augment the data-set and to create the synthetic large images is accessible on GitHub at https://github.com/mihaiboldeanu/POMO _Pollen_Classification. The data-set of cropped pollen images will be available on request to the authors.

ACKNOWLEDGMENT

(Mihai Boldeanu and Mónica González-Alonso contributed equally to this work.)

REFERENCES

[1] G. Burbach *et al.*, "GA^{2}LEN skin test study II: Clinical relevance of inhalant allergen sensitizations in Europe," *Allergy, Eur. J. Allergy Clin. Immunology*, vol. 64, no. 10, pp. 1507–1515, Oct. 2009.

[2] S. Adamov, N. Lemonis, B. Clot, B. Crouzy, R. Gehrig, M.-J. Graber, C. Sallin, and F. Tummon, "On the measurement uncertainty of Hirst-type volumetric pollen and spore samplers," *Aerobiologia*, Sep. 2021, doi: 10.1007/s10453-021-09724-5.

[3] C. Dananché, M.-P. Gustin, P. Cassier, S. T. Loeffert, M. Thibaudon, T. Bénet, and P. Vanhems, "Evaluation of hirst-type spore trap to monitor environmental fungal load in hospital," *PLoS ONE*, vol. 12, no. 5, May 2017, Art. no. e0177263. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28486534

[4] J. Oteros, A. Weber, S. Kutzora, J. Rojo, S. Heinze, C. Herr, R. Gebauer, C. B. Schmidt-Weber, and J. T. M. Buters, "An operational robotic pollen monitoring network based on automatic image recognition," *Environ. Res.*, vol. 191, Dec. 2020, Art. no. 110031. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0013935120309282

[5] J. R. Flenley, *The Problem of Pollen Recognition* (Problems in Picture Interpretation), M. B. Clowes and J. P. Penny, Eds. Canberra, ACT, Australia: CSIRO, 1968, pp. 141–145.

[6] R. R. Robbins, D. B. Dickinson, and A. M. Rhodes, "Morphometric analysis of pollen from four species of ambrosia (Compositae)," *Amer. J. Botany*, vol. 66, no. 5, pp. 538–545, May 1979. [Online]. Available: http://www.jstor.org/stable/2442503
[7] G. Mirkin and L. Bagdasaryan, "The feasibility of identifying paleontological objects with the aid of optical analysing systems," *Paleontol. J.*, vol. 6, pp. 103–108, Jan. 1972.
[8] M. Langford, G. E. Taylor, and J. R. Flenley, "Computerized identification of pollen grains by texture analysis," *Rev. Palaeobotany Palynol.*, vol. 64, nos. 1–4, pp. 197–203, Oct. 1990. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0034666790901334
[9] E. L. Vezey, V. P. Shah, and J. J. Skvarla, "A numerical approach to pollen sculpture terminology," *Plant Systematics Evol.*, vol. 181, nos. 3–4, pp. 245–254, 1992. [Online]. Available: http://www.jstor.org/stable/23674744
[10] A. Boucher, P. J. Hidalgo, M. Thonnat, J. Belmonte, C. Galan, P. Bonton, and R. Tomczak, "Development of a semi-automatic system for pollen recognition," Aerobiologia, vol. 18, nos. 3–4, pp. 195–201, 2002, doi: 10.1023/A:1021322813565.
[11] M. P. De Sá-otero, A. González, M. Rodríguez-Damián, and E. Cernadas, "Computer-aided identification of allergenic species of urticaceae pollen," *Grana*, vol. 43, no. 4, pp. 224–230, Dec. 2004, doi: 10.1080/00173130410000749.
[12] C. M. Costa and S. Yang, "Counting pollen grains using readily available, free image processing and analysis software," *Ann. Botany*, vol. 104, no. 5, pp. 1005–1010, Oct. 2009, doi: 10.1093/aob/mcp186.
[13] V. R. Dhawale, J. A. Tidke, and S. V. Dudul, "Neural network based classification of pollen grains," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Aug. 2013, pp. 79–84.
[14] A. B. Gonçalves, J. S. Souza, G. G. D. Silva, M. P. Cereda, A. Pott, M. H. Naka, and H. Pistori, "Feature extraction and machine learning for the classification of Brazilian savannah pollen grains," *PLoS ONE*, vol. 11, no. 6, pp. 1–20, Jun. 2016, doi: 10.1371/journal.pone.0157044.
[15] G. Astolfi, A. B. Gonçalves, G. V. Menezes, F. S. B. Borges, A. C. M. N. Astolfi, E. T. Matsubara, M. Alvarez, and H. Pistori, "POLLEN73S: An image dataset for pollen grains classification," *Ecolog. Informat.*, vol. 60, Nov. 2020, Art. no. 101165. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574954120301151
[16] S. Battiato, A. Ortis, F. Trenta, L. Ascari, M. Politi, and C. Siniscalco, "Pollen13K: A large scale microscope pollen grain image dataset," 2020, *arXiv:2007.04690*.
[17] J. Schiele, F. Rabe, M. Schmitt, M. Glaser, F. Haring, J. O. Brunner, B. Bauer, B. Schuller, C. Traidl-Hoffmann, and A. Damialis, "Automated classification of airborne pollen using neural networks," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 4474–4478.
[18] J. Schaefer, M. Milling, B. W. Schuller, B. Bauer, J. O. Brunner, C. Traidl-Hoffmann, and A. Damialis, "Towards automatic airborne pollen monitoring: From commercial devices to operational by mitigating class-imbalance in a deep learning approach," *Sci. Total Environ.*, vol. 796, Nov. 2021, Art. no. 148932. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969721040043
[19] O. Ronneberger, E. Schultz, and H. Burkhardt, "Automated pollen recognition using 3D volume images from fluorescence microscopy," *Aerobiologia*, vol. 18, no. 2, pp. 107–115, Jun. 2002.
[20] M. K. Sobol and S. A. Finkelstein, "Predictive pollen-based biome modeling using machine learning," *PLoS ONE*, vol. 13, no. 8, Aug. 2018, Art. no. e0202214. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30138366
[21] N. R. Nguyen, M. Donalson-Matasci, and M. C. Shin, "Improving pollen classification with less training effort," in *Proc. IEEE Workshop Appl. Comput. Vis. (WACV)*, Jan. 2013, pp. 421–426.
[22] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, no. 1, Jun. 2005, pp. 886–893.
[23] D.-C. He and L. Wang, "Texture unit, texture spectrum, and texture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 4, pp. 509–512, Jul. 1990.
[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.
[27] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.
[28] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
[29] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241.
[30] K. R. J. Oskal, M. Risdal, E. A. M. Janssen, E. S. Undersrud, and T. O. Gulsrud, "A U-Net based approach to epidermal tissue segmentation in whole slide histopathological images," *Social Netw. Appl. Sci.*, vol. 1, no. 7, p. 672, Jun. 2019, doi: 10.1007/S42452-019-0694-y.
[31] F. Long, "Microscopy cell nuclei segmentation with enhanced U-Net," *BMC Bioinf.*, vol. 21, no. 1, p. 8, Jan. 2020, doi: 10.1186/S12859-019-3332-1.
[32] A. Rakhlin, A. Davydow, and S. Nikolenko, "Land cover classification from satellite imagery with U-Net and Lovász-softmax loss," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 262–266.
[33] L. V. Stanislawski, E. J. Shavers, S. Wang, Z. Jiang, E. L. Usery, E. Moak, A. Duffy, and J. Schott, "Extensibility of U-Net neural network model for hydrographic feature extraction and implications for hydrologic modeling," *Remote Sens.*, vol. 13, no. 12, p. 2368, Jun. 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/12/2368
[34] J. Wu, Y. Zhang, K. Wang, and X. Tang, "Skip connection U-Net for white matter hyperintensities segmentation from MRI," *IEEE Access*, vol. 7, pp. 155194–155202, 2019.
[35] Y. Zhang, J. Wu, W. Chen, Y. Liu, J. Lyu, H. Shi, Y. Chen, E. X. Wu, and X. Tang, "Fully automatic white matter hyperintensity segmentation using U-Net and skip connection," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 974–977.
[36] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.
[37] J. Nagi, F. Ducatelle, G. A. D. Caro, D. Ciresan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, and L. M. Gambardella, "Max-pooling convolutional neural networks for vision-based hand gesture recognition," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Nov. 2011, pp. 342–347.
[38] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*.
[39] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, vol. 37, Jul. 2015, pp. 448–456.
[40] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[41] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*.

MIHAI BOLDEANU was born in Focşani, Romania, in 1992. He received the B.S. and M.S. degrees in electronic engineering and telecom from the Politehnica University of Bucharest (UPB), Romania, in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree in machine learning.

From 2017 to 2020, he worked as a Research Assistant with the National Institute for Research and Development in Optoelectronics and was a Researcher, from 2020 to 2021. Currently, he is working as a Data Science Researcher in the industrial/IoT sector. His publications are in aerobiology, atmospheric sciences, aerosol particle transport, bioclimatology, climate change, and machine learning.

MÓNICA GONZÁLEZ-ALONSO (Student Member, IEEE) was born in Pamplona, Spain, in 1991. She received the degree in biology, the Diploma degree in Research Training Program, and the M.Sc. degree in biodiversity, landscape, and sustainable development from the University of Navarra, Pamplona, in 2013 and 2014, respectively, where she is currently pursuing the Ph.D. degree in natural and applied sciences with the Environmental Biology Department, through the grant of ''Asociación Amigos Universidad de Navarra.''

She has worked with the research group Biodiversity Data Analytics and Environmental Quality, University of Navarra, since 2018. From June 2020 to December 2020, she was a Visiting Researcher with the Center for Allergy and Environment (ZAUM), Technical University of Munich, Germany. She accomplished the curse on ''Advanced Aerobiology Course 2019'' with a grant of the International Association for Aerobiology. Her publications are in aerobiology and ecology. She belongs to the Spanish Aerobiology Network and the European Academy of Allergy and Clinical Immunology.

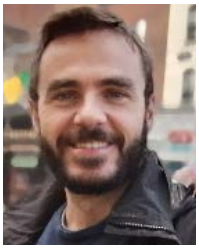
HORIA CUCU (Member, IEEE) was born in Braşov, Romania, in 1984. He received the B.S. and M.S. degrees in applied electronics and the Ph.D. degree in electronics and telecom from the Politehnica University of Bucharest (UPB), Romania, in 2008 and 2011, respectively.

From 2010 to 2017, he was a Teaching Assistant and then a Lecturer with UPB, where he currently serves as an Associate Professor. In this position, he has authored over 75 scientific papers in international conferences and journals, served as the project director for seven research projects and contributed as a researcher to ten other research grants. He holds two patents. He founded and leads Zevo Technology, a speech start-up dedicated to integrating state-of-the-art speech technologies in various commercial applications. His research interests include machine/deep learning and artificial intelligence, with a special focus on automatic speech and speaker recognition, text-to-speech synthesis, and speech emotion recognition.

Dr. Cucu was awarded the Romanian Academy Prize ''Mihail Drăgănescu,'' in 2016, for outstanding research contributions in spoken language technology, after developing the first large-vocabulary automatic speech recognition systems for the Romanian language.

CORNELIU BURILEANU (Life Senior Member, IEEE) received the Ph.D. degree in electronics and telecommunications from the Politehnica University of Bucharest, in 1986.

He has been a Professor with the Politehnica University of Bucharest, since 1995. He has participated in or led over 15 research grants; holds five national patents; and has published over 150 books, book chapters, and articles. His research interests include speech and voice recognition, machine learning, vocal signal analysis and coding, spoken terms recognition, and natural language processing.



JOSE MARIA MAYA-MANZANO was born in Badajoz, Spain, in 1983. He received the degree in environmental sciences and the M.Sc. degree in biological research from the University of Extremadura, Badajoz, in 2009 and 2010, respectively, the Ph.D. degree from the Vegetal Biology, Ecology and Earth Sciences Department, University of Extremadura, in 2015, and the M.Sc. degree in statistics from the National University of Distance Education (UNED), Madrid, Spain, in 2017.

He has worked in research groups of Spain, the U.K., Ireland, and Germany, from 2011 to 2022. From March 2018 to March 2020, he was the Project Manager in a project funded by the Irish Environmental Protection Agency (EPA), entitled Pollen Monitoring and Modelling (POMMEL). Since March 2020, he has been working as a Postdoctoral Researcher with the Center for Allergie and Environment (ZAUM), Technical University of Munich, Germany, working for the electronic Bavarian pollen monitoring networks. His publications are in aerobiology, modeling, allergens, climate change and time series analysis, species distribution models, and botany.

Dr. Maya-Manzano belongs to the Spanish Aerobiology Society; the Spanish Speakers Palynology Association; the European Aerobiology Society (being part of the elected committee); and the Association of Spanish Researchers, Germany.

JEROEN TITUS MARIA BUTERS was born in Eindhoven, The Netherlands, in 1957. He received the M.S. degree from the University of Groningen, The Netherlands, in 1984, and the Ph.D. degree from the University of Bern, Switzerland, in 1989.

He is the Deputy Director with the Center for Allergy and Environment, Technische Universität München. His publications are in aerobiology, allergies, environment, experimental toxicology with focus on the environment, xenobiotica metabolism, and skin.