Received May 27, 2021, accepted June 7, 2021, date of publication June 15, 2021, date of current version June 28, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3089538

An Intelligent Monitoring System for Assessing Bee Hive Health

DIOGO BRAGA[1,2], ANA MADUREIRA[1,2,3], (Senior Member, IEEE), FABIO SCOTTI[4], (Senior Member, IEEE), VINCENZO PIURI[4], (Fellow, IEEE), AND AJITH ABRAHAM[5], (Senior Member, IEEE)

[1]Interdisciplinary Studies Research Center, (ISEP/IPP), 4200-072 Porto, Portugal
[2]Institute of Engineering from Polytechnic of Porto (ISEP/IPP), 4249-015 Porto, Portugal
[3]INOV Institute for Systems and Computer Engineering, Technology and Science, 1000-029 Porto, Portugal
[4]Department of Computer Science, University of Milan, 20122 Milano, Italy
[5]Machine Intelligence Research Labs, Auburn, WA 98071, USA

Corresponding author: Vincenzo Piuri (vincenzo.piuri@unimi.it)

ABSTRACT Up to one third of the global food production depends on the pollination of honey bees, making them vital. This study defines a methodology to create a bee hive health monitoring system through image processing techniques. The approach consists of two models, where one performs the detection of bees in an image and the other classifies the detected bee's health. The main contribution of the defined methodology is the increased efficacy of the models, whilst maintaining the same efficiency found in the state of the art. Two databases were used to create models based on Convolutional Neural Network (CNN). The best results consist of 95% accuracy for health classification of a bee and 82% accuracy in detecting the presence of bees in an image, higher than those found in the state-of-the-art.

INDEX TERMS Bee monitoring, convolutional neural network, deep learning.

I. INTRODUCTION

Honey bees (i.e. western honey bees Apis mellifera) are the world's most frequent pollinators of natural ecosystems, averaging 13% of all floral visits. Furthermore, around 5% of plant species worldwide are exclusively visited by honey bees [1], with 200 economically important plants that require bee pollination for reproduction. In countries where agricultural production is an active role, honey bees are accepted as an important factor in modern agriculture, since they thrive in diverse climates, are domesticated, and can be manipulated by people. Related academic work indicates that crop quality and quantity can be increased by using honey bees to facilitate crop pollination, even in self-pollinating plant species, with some crops' yields depending 100% on honey bees and other pollinator insects [1], [2]. As stated by the UN Food and Agriculture Organization [3], more than 75% of the world's food crops rely to some extent on pollination for yield and quality. Without pollinators, apples, coffee, tomatoes, and cocoa, among others, would not exist. The European Commission highlights that pollinators provide vital ecosystem services to crops and wild plants, forming a key component of European biodiversity [4].

However, the honey bee species should be constantly monitored to prevent problems and diseases. For example, Colony Collapse Disorder (CCD) is a phenomenon characterized by an unexplained rapid loss (60-90%) of a colony's adult population [5], [6]. One odd characteristic of CCD colonies is that they usually have plenty of food stores and can even contain a queen and a small number of young workers, and the colony's reserves remain untouched by robbing bees or honey bee comb pests for several weeks after the collapse. Furthermore, CCD colonies rarely have the bodies of the dead bees in the hive [5]–[7]. Some of the potential causes for CCD consist in parasites (e.g., varroa destructor, acaparis woodi), insecticides, genetic mutations, climate conditions and viruses and fungi [7]. Fig. 1 shows that in the United States, the total winter loss of bees has been exceeding the acceptable loss since 2009, and is expected to keep this trend [8]. Furthermore, small cases of CCD still exist in specific regions and apiaries [6].

Forest fires, human induced stress, poor nutrition, pollution, biodiversity loss and intensive agriculture still pose as threats for the survival of honey bees [3], [9], [10]. Human induced stress can occur when beekeepers tend to their bee

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

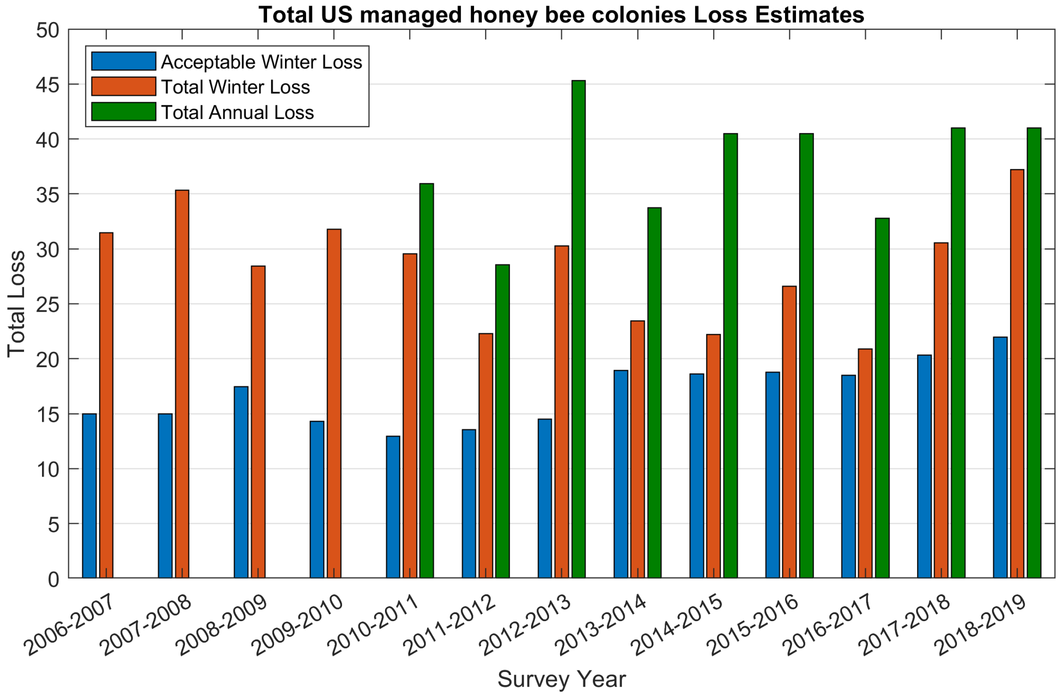

FIGURE 1. US honey bee losses between 2006 and 2019 [8].

hives, as it is often interpreted as an intrusion by the colony's bees. Furthermore, from interviews with companies that work in the area, a beekeeper might be required to inspect up to 100 hives daily, making the detection of health problems in each hive have a short amount of time.

The world population is expected to rise in the upcoming years, creating a demand for higher food production to sustain this growth. The western honey bees pollinate crop species that compose up to one third of the average diet, worldwide [11]. Considering this expected growth and the current methodology of hive health detection, there is a demand for the development of solutions that can aid preserving honey bees in a healthy state, since these play a vital role in the produces provided to humans by ecosystem services.

On the context of hive health classification, it is important to state how a hive's health can be assessed. From literature, it can be stated that the health of a population can be obtained by aggregating the health of its individuals (i.e. "rolling up" the individual level data), which results in a summary statistic that can be used to measure the population health [12]. However, surveying or analysing an entire population of individuals can often be impossible or impractical, requiring that the analysis is conducted only on a sample of the population [13]. As such, in order to assess the health status of a bee hive (the population), one must be able to first assess the health of the bees (individuals). Advances in Deep Learning (DL) and Computer Vision (CV) have shown potential applications in the automatic detection of bees and their respective health, such as detecting environment conditions (i.e., temperature, humidity, etc.), detecting signs of varroa destructor, among others [11], [14]–[18]. Considering these advancements, the classification of a bee's health status can be performed through images of the specimen, using CV and DL. However, in order to obtain images of bees, there must also exist a previous task that can detect images of bees in a scene and automatically crop the detected bees into new images.

This work aims to develop a system to solve a class of applications by monitoring the health status of honey bee hives using CV and DL techniques. The proposed system can decrease the need for human intervention in the hive, lowering induced stress in bees and increasing expected yield. Furthermore, considering that treatment effectiveness of hives with diseases, and neighboring healthy hives, can be increased depending on whether the treatment is administered at an early stage, it is also important to have a system that can aid with an early detection of health problems. The work done in [21], made available on Kaggle, was used as the basis for this study. The novelty of the methodology proposed in this paper consists in its high bee health classification accuracy when compared to those found in the state of the art, whilst maintaining the same efficiency.

On the context of the goals of this study, the following tasks for the development of a hive health classification system were identified:

- *Object Detection*: develop an algorithm to detect the presence of honey bees in images. The algorithm will detect and crop any honey bees detected in an image that have the minimum required quality to be later analysed.

- *Health Classification*: develop an algorithm to classify the health status of a bee hive, using cropped images of honey bees. Considering the work developed in [21], the approach defined for this task will consist in improving the classification results obtained by the authors, using hyper-parameter optimization, and the application of image filters and mathematical morphology operators, separately. The health status of the bee will be used to assess the overall health of the colony, as previously mentioned.

The remainder of this paper is structured as follows: Section II provides a brief overview of available literature and concepts used in this work; Section III describes the data used for the purposes of this study; Section IV specifies the methodology defined in this study, which also details the proposed system architecture design; Section V details the experimental results of the stated trials, along with respective results discussion; Section VI summarizes the major findings of this work.

II. LITERATURE REVIEW

As stated above, contributions have been made in DL and CV, which show potential applications in this study's scope. Furthermore, work has been made regarding DL and bees. This section aims to provide a review of these contributions.

In the last decade, advances have been made to the CNN architecture, depending on the goal of its application (i.e., image classification, object detection, image segmentation). Some of the most significant contributions are [22]: AlexNet, which won the 2012 ILSVRC competition (one of the most difficult challenges for image detection and classification, at the time); GoogleNet, which won the 2014 ILSVRC competition (introduced the concept of split, transform and merge blocks); ResNet, proposed by Microsoft, for the net training of 150 layers deep networks; and DenseNet, which uses the idea of cross channel connectivity. One of the most significant contributions is Mask-RCNN, which is a Recurrent Convolutional Neural Network that extends Faster R-CNN, which is one of the best architectures for object detection and image segmentation [23]. Mask-RCNN implements a branch for predicting an object mask in parallel with the branch that performs bounding box recognition. With this architecture, Mask-RCNN can perform faster and simpler training, with only a 5 Frames Per Second (FPS) loss [23].

In the literature, recent contributions have been made involving Machine Learning (ML) algorithms and bee health monitoring. The work done by Kulyukin *et al.* [15] used 9110 audio samples, equally distributed by "Bee Sound", "Noise Sound" and "Cricket Sound", in order to monitor a bee hive. The approach used a CNN based architecture, and was tested on the BUZZ1 and BUZZ2 [24] data sets against other types of ML & DL algorithms. The study concluded that DL can be used to monitor bees in a bee hive, with the CNN based approach having obtained an accuracy of 95.21% and 96.53% on the BUZZ1 and BUZZ2 data sets, respectively. The study conducted in [18] analysed standard ML techniques (Random Forest, SVM, KNN and Logistic Regression) to perform classification of bee audio between 3 classes (bee, noise and cricket). The work concluded that ML based techniques can aid in the classification of bees using audio, as well as help monitoring a bee hive's health. In [16], the authors compared DL (DCNN) and ML (SVM with linear, RBF and 3rd order polynomial kernels) approaches for bee hive sound recognition. The study used an annotated data set of 78 recordings, comprising approximately 12 hours of audio. From the compared approaches, the study concluded that the SVM outperformed CNN. However, the authors also recognized that CNN showed room for improvement, highlighting the impact that samples with large context and the size of training data have on CNN. The authors in [25] used video and CNN for the automatic detection of honey bees in a hive, using approximately 11 hours of video footage, where bees have been placed with a tag. The best F1-score value achieved in the study was of 0.686 and the authors concluded that the manual labeling provided by the tags may not suffice for bee detection. F1-score is a metric to evaluate models that corresponds to the harmonic mean between the precision and recall metrics [26]. Both [14]– [17] used CNN and ML algorithms to perform bee hive monitoring, aiming at the detection of mites and varroa destructor. The first study obtained an accuracy of 93% detection of varroa destructor using a training data set of 5000 artificially generated images, tested with different CNN configurations [14]. The second study used different configurations of lights in a special camera setup, using 1920×1080 images recorded at 50-60 FPS. The authors highlighted the challenge in detecting mites on moving bees or on bees where wings occlude the mites [17]. The work done in [21] provides a thorough explanation and exploration of the data set publicly available in [27]. The approach consists of 2 CNN models to perform the classification of a bee's health and subspecies. The author of [21] provides several methods to analyse and remove bias from the data set, as well as a solid pipeline for the modular training of the CNN models. Furthermore, using the data set provided in [27], the author was able to obtain a best accuracy of 84.92% and 86.54% on the health and subspecies models, respectively. Due to the solid pipeline and results presented in [21], the author's work will be used as a baseline for the image classification task of this study, as well as a comparison for obtained results. In the Kaggle kernel of [21], the author implemented a basic pipeline to load the data set in an efficient manner and to train classifiers using a balanced data set. When any data set is loaded, the pipeline checks how many instances belong to each class and proceeds to balance the data by splitting in accordance with class distribution, forcing all classes to have the same distribution, minimizing bias. Additionally, the pipeline also implements an image data generator (i.e., the standard Keras ImageDataGenerator) which randomizes image attributes by applying operations such as scaling, skewing, translation, flipping and zooming, among others. The baseline work's learning algorithm is DCNN. In the baseline work, the author

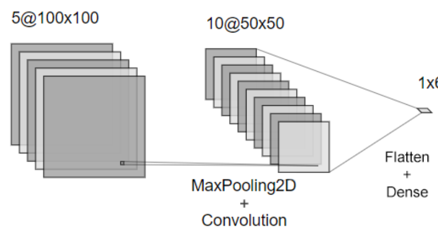

FIGURE 2. DCNN architecture of the baseline approach for the health classification of bees.

TABLE 1. Possible values of each attribute of the bee image health classification data set.

Attribute	Values
Health	Varroa, Small Hive Beetles Ant Problems Few Varroa, Hive Beetles Healthy Hive Being Robbed Missing Queen
Pollen Carrying	True False

also created methods to visualize the kernel matrices produced by developed DCNN, as well as developed a few classifier models for the health, subspecies and pollen_carrying attributes, which will be used in this study for comparison of the results obtained. All images are scaled to have the same size (i.e., 100·100 pixels). Fig. 2 showcases the DCNN architecture used in [21] for the health classification, which is similar to the architectures used for the subspecies and pollen_carrying classifiers.

This paper's proposed approach is able to obtain a higher efficacy in respect to the work described in [21], whilst maintaining similar efficiency. Furthermore, it can also be stated that this work adds onto the work conducted in [21], as an object detection model for the automatic image cropping of bees and subsequent health classification was developed, so that the task of classifying a hive's health can be fully automated.

III. DATA DESCRIPTION

For this study, different data sets were required for each problem task. The goal of this section is to provide a detailed description about the data used for each problem task, including decisions made about labels used for the bee health classification task and the conditions related to the data collection process of the bee object detection.

A. BEE HEALTH CLASSIFICATION

Considering the task of bee health classification through image, the data set used in this study is from [27]. The data corresponds to a set of cropped honey bee images, each annotated with several attributes. Of these, the health, subspecies, and pollen_carrying are particularly relevant to this study, as they correspond, respectively, to the current state of the bee hive's health, the subspecies of the honey bee and whether it is carrying pollen or not. Considering these attributes, *health* was considered for the development of a classifier, as it directly reflects the health status of the corresponding bee. Furthermore, *subspecies* and *pollen_carrying* were also considered for separate classifiers, as they can be used to further determine health problems in a bee hive. On the case of the pollen_carrying attribute, counting the number of bees with pollen and analysing shifts in this count can help determine if a hive has food supplies, whilst the subspecies can determine if hive robbing is occurring. Hive robbing occurs when the bees of a different hive (which can contain a separate species of bees) invade another hive with the goal of collecting nectar and honey [28]. From these examples, the potential of the attributes in bee health classification can be identified.

Table 1 shows the different values (classes) that each attribute can have on the context of the data set's problem. As can be seen, some of the attributes (like subspecies) account for unknown examples. The health attribute was annotated by experts in the field, when the image of the honey bee was taken. Furthermore, the health attribute also contains some problems related to the hive (such as ''Ant Problems'') as the state of bee hives can reflect changes in honey bees and vice versa [29].

The data set is composed of 5,172 RGB images of honey bees, with varying sizes (width between 27 and 392 pixels; height between 24 and 520 pixels). As this study's work will use the baseline in [21], the data preparation methods developed in the baseline were also used. Specifically, the data is split in accordance with class distribution, and an image data generator is used to randomize images by applying scaling, skewing, translation, flipping and zooming operators. Furthermore, the images were all scaled to have the same size (i.e., 100·100 pixels).

B. BEE OBJECT DETECTION

Considering the task of developing the object detection model to automatically crop images of bees, the data set was extracted from Explore's honey bee live streams (i.e., [30] and [31]). The first is an RGB live stream from the perspective of the landing pad of two different apiaries. The perspective and angle of the camera is distinct in each apiary and provides clear sight of the bees that are landing, as well as those immediately outside of the apiary, as can be seen in Fig. 3, with two perspectives of honey bees' landing. Fig. 4 shows the second live stream of Grayscale images bees inside an apiary, from a top-down perspective.

A total of 2,750 honey bee footage images were extracted from the live streams (at a rate of 5 frames per second). Of these images, 1,533 and 1,217 correspond to the first and second live streams, respectively. All images were extracted with an image size of 990·504 pixels. For this data, the Gold Standard (manual) was chosen for the annotation process (versus an automated approach, i.e., Silver Standard), as the quality of the annotations produced by the

FIGURE 3. Example of footage from the first live stream of honey bees.

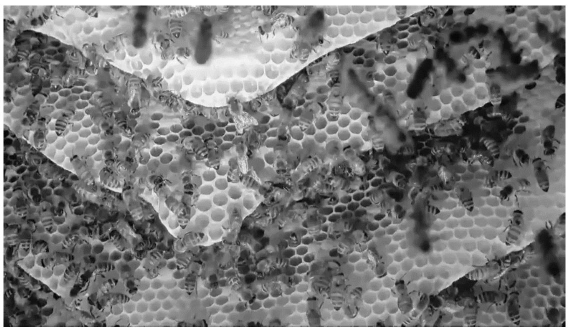

FIGURE 4. Example of footage from the second live stream of honey bees.

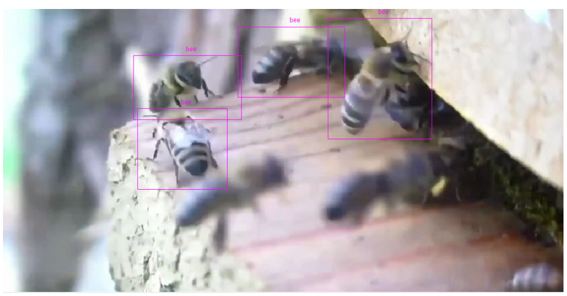

FIGURE 5. Example of annotations for object detection.

Gold Standard are usually of higher quality than those of the Silver Standard, which can affect the quality of the developed model [32]. During manual annotation, images that were too blurry were not annotated. As such, with the Gold Standard high quality image annotation can be obtained, despite a potential decreased number of samples, when considering the same annotation timespan. The images were manually annotated in accordance with the following criteria: only bees that are at least 50% visible and whose blur is minimal are annotated for object detection. As such, Fig. 5 showcases an example of the annotations made following the previously mentioned criteria, using a custom-made tool for annotation. As can be seen in Fig. 5, bees whose quality is too low (e.g., blurry) are not annotated.

The Gold Standard annotation produced 251 images, which were used for the object detection phase. Of these images, 32 correspond to empty bee hive, 17 contain flowers with no bees and the remaining are in respect to the second data set. The empty images have the goal to further refine the model with examples of scenes where bees normally are present. Due to computational restrictions, all images and respective annotations were scaled down to 576·290 pixels.

IV. PROPOSED METHODOLOGY

Considering the goals set for the work conducted in this study, it can be stated that the methodology branches into two different classes of methods: health classification through images of bees (where the object in focus is a bee), and automatic detection and cropping of bees in images (so that these images can be provided to the health classification model). The remainder of this section details the bee health classification and object detection methods, each in their respective subsection.

A. BEE HEALTH CLASSIFICATION

The proposed methodology for the health classification of bees follows a computational schema to the one developed in [21]. As stated previously, considering the goal of improving the results obtained by the authors of [21], the following approaches can be defined:

- Application of image filters and mathematical morphology operators, such as opening and closing.
- Hyper parameter optimization of the classifiers and respective architecture.

TABLE 2. Image filters and mathematical morphology operators used on the health classification task.

Image Filters	Mathematical Morphology Operators
Color Dodge, Color Burn, Hard Light	Opening, Closing, Dilation, Erosion

Table 2 shows the image filters and mathematical morphology operators used to improve the classification results. Both preprocessing techniques were tested independently of each other, but the possible values within each of the preprocessing techniques were combined, as to determine their effectiveness at improving the classification results. Fig. 6 and Fig. 7 exhibit the application of different image filters and

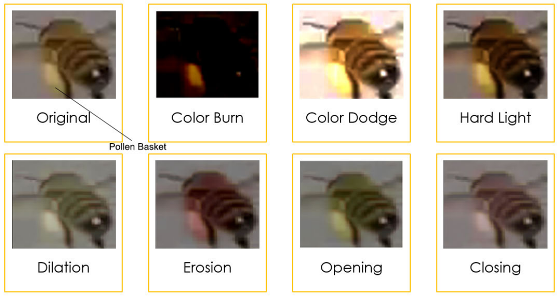

FIGURE 6. Application of image filters and mathematical morphology operators in a bee that is carrying pollen.

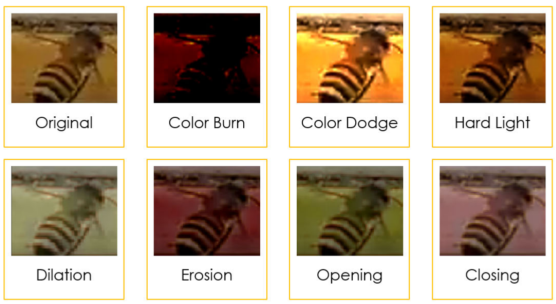

FIGURE 7. Application of image filters and mathematical morphology operators in a bee that is not carrying pollen.

mathematical morphology operators on a bee that is carrying pollen and another that is not, respectively.

Considering the case depicted in Fig. 6, some filters seem to brighten the region where the pollen is seen (near the bee's legs), such as the case of Color Dodge, whilst others seem to make it more difficult to identify any visual clues. On the case of Fig. 7, which corresponds to a bee not carrying pollen, some filters seem to highlight some information, whilst others seem to make it more difficult to identify visual clues.

Regarding the hyper parameter optimization, the goal is to improve the classification results by finding the best combination of values for the DCNN architecture of the baseline work, in a grid search like fashion. Table 3 summarizes the hyper parameters that were optimized in this study. Regarding the variation in the number of layers, preliminary testing showed that there were no significant improvements in increasing the number of convolutional layers past 4. Similarly, the same results were obtained regarding the number of neurons in each layer, as when compared to those in Table 3. Regarding other parameters' possible values, as the goal is to improve the results obtained from the baseline approach, performing an exhaustive search on a wide variety of values is important, as to ensure a desirable confidence on statements about the best configuration for the CNN. Furthermore, several random seeds were used in order to determine which initial configuration of the weights of the CNN provided better results, as well as to assess the overall improvement of the architectural changes to the CNN. The latter motive arises from the fact that if a change in the

TABLE 3. CNN hyper-parameters that were optimized in this study.

CNN Hyper-Parameters Optimized
Number of layers; Number of neurons in each layer; Random seed initial value; Pooling methods; Activation functions; Optimizers; Loss functions;

TABLE 4. Different possible values tested for each hyper parameter during the optimization trial.

Hyper-Parameter	Values
Structure	Layer 1: [2,5,10, 15] Layer 2: [10, 15, 20, 25, 30, 35] Layer 3 (optional): [10, 15, 20, 25, 30] Layer 4 (optional): [5, 10, 15]
Pooling Method	AveragePooling2D, MaxPooling2D
Optimization Function	Adam, Nadam, Adamax, Adadelta
Loss Function	Mean Squared Error, Categorical Cross Entropy, Mean Squared Logarithmic Error, Poisson, Log-cosh, Kullback Leibler divergence
Random Seed	42, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 150, 160, 270, 560, 1603, 6141, 410

architecture improves the obtained results, then it should do so in several random states, therefore reducing the possibility that an increase in classification results was due to a beneficial random state.

Furthermore, each of the values of the hyper-parameters were combined with the others in separate phases, in order to reduce the total number of combinations tested in one given test setup (which would total to around 516 thousand different runs). After each phase, values that produced significantly poor results were discarded for next setups. This process was repeated until all hyper parameters were stabilized (i.e., had obtained the best possible results). The last setup was composed of a total of 576 different tested combinations. Additionally, only the last setup used all random seed values, whereas the others used at least 3. Table 4 shows each of the values that were tested, discriminated by their respective hyper parameter.

Each model configuration, was run for 20 epochs, with 50 steps per epoch, as to match the same training efforts conducted on the baseline approach. The data set was split into train, test, and validation data, with a split weight of 0.65, 0.25 and 0.10, respectively. The validation set was used during training to provide continuous validation of the model, whilst the test set was used for the final validation. As mentioned previously, by using the same training conditions as those of the baseline approach, a further comparison can be made on the beneficial impacts of the defined approaches. Additionally, every time a new configuration of the CNN was executed, the image data generator was reset, as to ensure that any improvements of the classification results happen due to a change in the preprocessing and/or architecture of the CNN, instead of different quality images.

The accuracy, validation loss and FPS metrics were considered the most meaningful attributes to evaluate the data set. Since the data set was always balanced in accordance with class distribution, the accuracy metric can be used, as bias is minimal [33]. However, recall, precision and F1-score were also used during model validation, for confirmation of the accuracy scores. Validation loss was used as it empirically corresponds to a measure of how well a model was generalizing to unseen records [34]. Finally, the FPS was used to assess the speed at which the model can operate, as well as its fitness for a real-time system. All FPS results were obtained using the same measurement method, which consisted of executing the model's detection method as many times as possible, within 60 seconds. The result was then calculated by averaging the number of detections performed over the 60 second period.

B. BEE OBJECT DETECTION

The bee health classification task depends on images of cropped bees. Considering this dependency, and as stated previously, the goal of this task is to develop an object detection model that provides cropped images of bees. The model achieves this end using an image that may contain bees as input and obtaining the bounding box coordinates of any bees present in the image as output, allowing for automatic cropping of bee images that can be used by the bee health classifiers.

For this task, the proposed methodology consists in using a Mask-RCNN with its default configuration and specific altered hyper-parameters. Additionally, the Mask-RCNN algorithm used was trained from scratch, without transfer learning (i.e., no pre-trained weights were provided to the network).

TABLE 5. Computational environment details, discriminated by physical and virtual setup.

Hyper-Parameter	Values
Graphics Card	Nvidia Geforce GTX 1050 Ti 4GB
CPU	Intel i7 HexaCore 8700 3.20GHz
RAM	32GB
Language	Python
Environment	Keras with Tensorflow GPU backend; tensorflow v1.13.1; keras v2.2.4; mask-rcnn v2.1; cuda v10.0.130; cudnn v7.6.0;

For the bee object detection task, accuracy and average Intersection over Union (aIoU) were used as the evaluation metrics. The latter corresponds to the intersection of the predicted mask and ground truth mask, over the union of the two masks. The accuracy was measured using IoU, with a threshold of 0.5 (i.e., an IoU higher than 0.5 is considered a true positive). Table 5 shows the computational environment on which all trials and development efforts were conducted.

Regarding the object detection task, the cross validation technique chosen was the train/test split. Even though the data set is not as voluminous as the health classification task's data set, the chosen cross validation still provides reliable results, since, regarding object detection, cross-validation techniques provide minimal differences in accuracy [35]. Furthermore, due to resource constraints, other cross validation methods are impractical to use. The train/test split was of 0.7/0.3, and all splits of the data set were balanced in accordance with class distribution.

V. EXPERIMENTAL RESULTS & DISCUSSION

This section provides details about the experimental results obtained for the bee health classification and detection tasks, including constraints considered. Finally, it provides a discussion of the obtained results.

A. EXPERIMENTAL RESULTS

This subsection provides details about the approach used for both tasks, such as results obtained for each model, constraints considered, and optimizations performed.

1) BEE HEALTH CLASSIFICATION

As mentioned previously, the purpose of the health classification task is to determine the health, subspecies, and pollen_carrying values of a bee through image. In order to improve the results found in the baseline work [21], the following trials were conducted:

1) Use a combination of image filters and mathematical morphology operators to enhance features of the image, to improve the validation accuracy.
2) Optimize the hyper parameters used in the CNN to achieve the best validation accuracy, considering the optimal values for the architecture.

TABLE 6. Best results obtained using baseline work's optimal CNN.

Measure	Baseline (Best)		
	health	pollen_carrying	subspecies
Accuracy	0.8492	0.9574	0.8654
Precision	0.8200	0.9900	0.91000
Recall	0.8500	0.9600	0.8700
F1-Score	0.8100	0.9700	0.8800
Val. Loss	0.4052	0.1217	0.3319
FPS	642	737	773

TABLE 7. Best results obtained using image filters and morphology operators applied to the baseline work's optimal CNN.

Measure	Image Filters (Best)		
	health	pollen_carrying	subspecies
Accuracy	0.7092	0.9907	0.6798
Precision	0.7200	0.9900	0.69000
Recall	0.7100	0.9900	0.6800
F1-Score	0.7100	0.9900	0.6700
Val. Loss	1.1614	0.0661	1.3782
FPS	627	691	689
Operator	Closing	Opening	Opening

To compare any results obtained with the approaches defined in this study, the baseline work's (developed in [21] optimal CNN was run for each of the target attributes, as per Table 6. The best results obtained with hyper parameter optimization trial are available in Table 8. Table 7 shows the results obtained from the application of image filters and morphology operators, using the baseline CNN model best configuration present in [21]. Each filter was applied during the preprocessing phase when the image is loaded.

From the results shown in Table 8, it can be stated that these are significantly better than the ones proposed in the

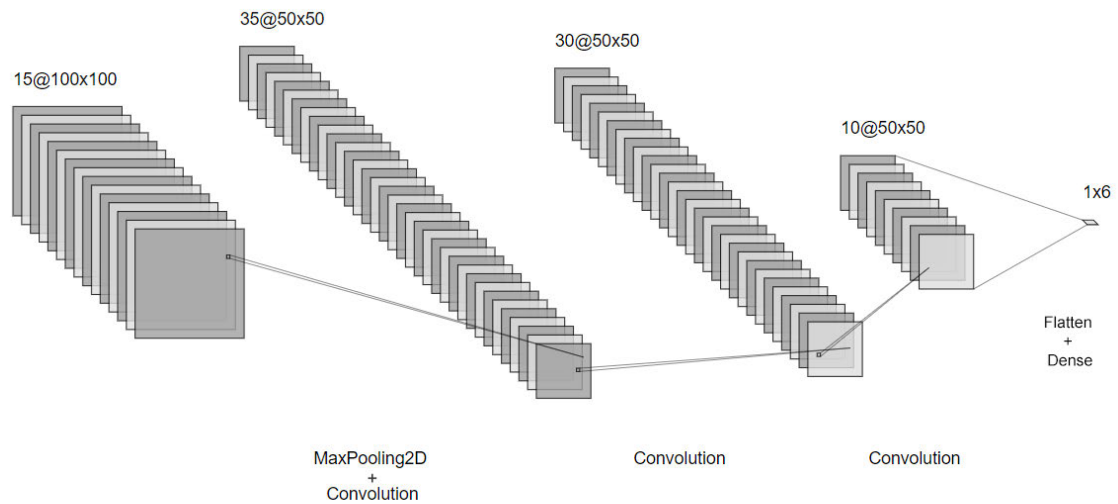

FIGURE 8. DCNN architecture of the hyper parameter optimized approach for the health classification of bees, developed in this study.

TABLE 8. Best results obtained with hyper-parameter optimized version of the baseline work's CNN.

Measure	Hyper Parameter Optimized (Best)		
	health	pollen_carrying	subspecies
Accuracy	0.9520	0.9822	0.9126
Precision	0.9500	0.9900	0.9200
Recall	0.9500	0.9800	0.9100
F1-Score	0.9500	0.9900	0.9200
Val. Loss	0.1505	0.4655	0.2420
FPS	568	734	770

base methodology. Table 9 shows the best architecture found for each of the attributes tested, whose results are summarized in Table 8. Fig. 8 showcases an example of the best hyper-parameters found for the health label converted into the respective CNN architecture. A similar architecture like the one found in Fig. 8 can be obtained for the case of subspecies attribute, using values [15, 25, 30, 15] for neurons and AveragePooling2D, and the same can be applied to the pollen_carrying attribute, using values [10, 20, 20, 5] and AveragePooling2D. All these values can be found in Table 9.

2) BEE OBJECT DETECTION

The goal for the object detection task is to detect bees in images, so that these can be cropped and analysed by the health classification task. To perform the object detection, the default configuration of Mask RCNN was used with the changes stated in Table 10 added to the training configuration. The NUM_CLASSES parameter was changed to reflect the number of classes for the detection (i.e., 'background' and 'bee'). The TRAIN_ROIS_PER_IMAGE and IMAGES_PER_GPU were reduced due to computational constraints, as to lower the amount of VRAM required to train the model. Furthermore, IMAGE_MAX_DIM and IMAGE_MIN_DIM reflect the minimum values for the dimensions of the images to be used in the model that are divisible by 64, which corresponds to a requirement of Mask RCNN.

TABLE 9. CNN best architecture for maximizing the classification results, discriminated by attribute.

Attribute Tested	Best Architecture	
health	Structure	[15, 35, 30, 10]
	Activation functions	[ReLU, ReLU, ReLU, softmax]
	Pooling method	MaxPooling2D
	Optimizer	adam
	Loss function	categorical cross-entropy
	Random seed	42
pollen_carrying	Structure	[10, 20, 20, 5]
	Activation functions	[ReLU, ReLU, ReLU, softmax]
	Pooling method	AveragePooling2D
	Optimizer	adam
	Loss function	categorical cross-entropy
	Random seed	42
subspecies	Structure	[15, 25, 30, 15]
	Activation functions	[ReLU, ReLU, ReLU, softmax]
	Pooling method	AveragePooling2D
	Optimizer	adam
	Loss function	categorical cross-entropy
	Random seed	100

TABLE 10. Mask RCNN configuration for training the object detection model.

Parameter	Value
NUM_CLASSES	2
IMAGE_MAX_DIM	576
IMAGE_MIN_DIM	320
GPU_COUNT	1
IMAGES_PER_GPU	1
TRAIN_ROIS_PER_IMAGE	32
STEPS_PER_EPOCH	128

For the object detection, each model was trained up to 100 epochs. Additionally, each training session's random seed was set to 800, as to have a reference for performance comparison.

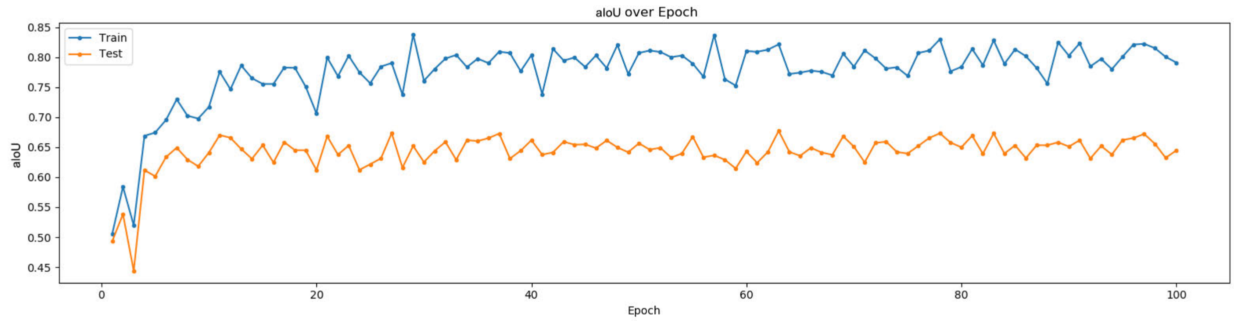

FIGURE 9. Train and test aIoU scores over each epoch.

TABLE 11. Best results obtained for the bee detection task using mask RCNN (epoch 063).

Data set	aIoU	Accuracy	FPS
Train	0.8213	0.8971	2.7
Test	0.6773	0.8289	2.7

From the results depicted in Table 11, it can be stated that the aIoU is very promising (0.6773), including the respective accuracy of the aIoU value (which is 82.89%), despite the number of images used for the training process. Furthermore, as can be seen in Fig. 9, there seems to be no significant increase in test aIoU scores past epoch 011, meaning that there is also no significant increase in accuracy. The same occurs for train aIoU scores after epoch 037. Regarding the FPS, it can be stated that it is significantly low. Even when tested with lighter configurations (resnet50 instead of resnet101; 320 $\times$ 256 images), the max FPS reached is 3. Additionally, Table 11 does not present any result comparison between state of the-art techniques, as it was not possible to find related relevant work for comparison.

B. RESULTS DISCUSSION

Comparing the results obtained in the health classification task, it can be stated that the approach defined in this study yielded significantly better results than those obtained in the baseline work [21], which corresponds to the best work available in the literature. The hyper parameter optimization was able to achieve significantly better results across several metrics, with highlight to the increase in accuracy. The health attribute yielded the most significant improvement in accuracy, as it increased by 10% (95%), when compared to the baseline work (85%). The attribute pollen_carrying had the least increase across all scores, having only increased its accuracy in approximately 2%. Furthermore, it was possible to obtain the architecture configuration for each of the best results obtained. However, the FPS of all developed models for the health classification have lowered. This result is expected, as the architectures obtained either contain additional preprocessing (i.e., image filters), or have a more complex structure (i.e., hyper parameter optimization). Nevertheless, the decrease in FPS is low enough not to have any significant impact in the viability of the monitoring system's real time operation capabilities. This can be verified by converting the FPS of the baseline and hyper-parameter optimized approaches for the health attribute to milliseconds, which equal respectively to 1.56 and 1.76. The difference between the times is 0.2 milliseconds, which can be discarded as significant in the context of this study.

Regarding the image filters and mathematical morphology operators, it can be concluded that there is a significant decrease across all scores when using these, except for the case of pollen_carrying, as there was an increase in accuracy when using the opening operator. These results suggest that image operators will only aid the classification process for the pollen_carrying and should not be used for any of the other attributes. This arises from the fact that, for example, the best accuracy obtained for the health attribute was 70%, which corresponds to 15% less than the one obtained with the baseline model.

An analysis of the obtained results in the object detection task concludes that the obtained accuracy is promising. Therefore, there is potential in integrating the approach defined for the object detection task with the health classification model. Regarding the FPS of the object detection approach, it can be stated that it is significantly lower than the value required for a real time system. As such, although the health classification model can be applied in a real time scenario, the addition of the model that performs object detection makes it unsuitable for such a scenario's requirements. However, it can be concluded that the efficiency of the model developed in this is as expected for the algorithm used (i.e. Mask RCNN), since the FPS results obtained for the object detection are similar to those found in [23] (when also considering differences in hardware). However, it is also possible to state that a better computational environment, specifically hardware, might make the object detection's integration feasible.

VI. CONCLUSION

In sum, it can be stated that honey bees have a vital role in human activities worldwide. From the literature review, it can be concluded that there is potential for the application of CV and DL to aid the survival of honey bees. Furthermore, the creation of automatic bee monitoring techniques can greatly benefit the health of honey bees and global food production. As such, it is expected that the system can perform detection with high frequency, nearing a real-time system.

The approach developed for the health classification of honey bee images was able to obtain significantly better results than those obtained by the baseline methodology (i.e., the one defined in [21]). The developed approaches' FPS is not significantly high for the overall methodology to be feasibly implemented in a real time system-like environment. Although the health classification task yielded very high FPS for each classifier (over 500FPS), the object detection approach runs at 2 to 3 FPS, slowing the health classification task's speed. However, this efficiency is expected when using Mask RCNN [23], and may be able to be increased using better hardware. The health classification model's results show great potential for the reduction of human induced stress associated with the monitoring of honey bees, being feasible to implement this model with other approaches.

The next steps will consist of improving both the detection accuracy and speed of the proposed model, as well as applying the system in a practical scenario, to assess any needs and adjustments for the successful implementation of the bee hive health monitoring system. Additionally, Faster R-CNN will be considered for searching a classifier able to achieve similar accuracy, but with a faster computational time with respect to the one presented in this work.

ACKNOWLEDGMENT

This study was developed under the 2nd Edition of the New Talents in Artificial Intelligence Program of the Calouste Gulbenkian Foundation, with the partial support of the University of Milan's APC initiative.

REFERENCES

[1] K.-L.-J. Hung, J. M. Kingston, M. Albrecht, D. A. Holway, and J. R. Kohn, "The worldwide importance of honey bees as pollinators in natural habitats," *Proc. Roy. Soc. B, Biol. Sci.*, vol. 285, no. 1870, Jan. 2018, Art. no. 20172140, doi: 10.1098/rspb.2017.2140.

[2] C. E. Aslan, C. T. Liang, B. Galindo, H. Kimberly, and W. Topete, "The role of honey bees as pollinators in natural areas," *Natural Areas J.*, vol. 36, no. 4, pp. 478–488, Oct. 2016, doi: 10.3375/043.036.0413.

[3] Food and Agriculture Organization of the United Nations. (May 20, 2018). *FAO—News Article: Bees Must be Protected for the Future of Our Food*. Accessed: Oct. 21, 2020. [Online]. Available: http://www.fao.org/news/story/en/item/1132329/icode/

[4] European Commission. (Jan. 2015). *Final Report Summary—STEP (Status and Trends of European Pollinators) | Report Summary | STEP | FP7*. Accessed: Oct. 21, 2020. [Online]. Available: https://cordis.europa.eu/result/rcn/176061_en.html

[5] R. M. Underwood and D. vanEngelsdorp, "Colony collapse disorder: Have we seen this before?" *Bee Culture*, vol. 135, no. 7, pp. 13–18, 2007.

[6] D. vanEngelsdorp, K. S. Traynor, M. Andree, E. M. Lichtenberg, Y. Chen, C. Saegerman, and D. L. Cox-Foster, "Colony collapse disorder (CCD) and bee age impact honey bee pathophysiology," *PLoS ONE*, vol. 12, no. 7, Jul. 2017, Art. no. e0179535, doi: 10.1371/journal.pone.0179535.

[7] A. Nath. (2016). *Colony Collapse Disorder*. Accessed: Nov. 21, 2018. [Online]. Available: https://www.researchgate.net/publication/307989119_Colony_Collapse _Disorder

[8] Bee Informed. (Jul. 26, 2019). *Honey Bee Colony Losses 2018-2019—Preliminary Results | Bee Culture*. Accessed: May 3, 2021. [Online]. Available: https://www.beeculture.com/honey-bee-colony-losses-2018-2019-preliminary-results/

[9] A. Jacques, M. Laurent, M. Ribière-Chabert, M. Saussac, S. Bougeard, G. E. Budge, P. Hendrikx, M.-P. Chauzat, and E. Consortium, "A pan-European epidemiological study reveals honey bee colony survival depends on beekeeper education and disease control," *PLoS ONE*, vol. 12, no. 3, Mar. 2017, Art. no. e0172591, doi: 10.1371/journal.pone.0172591.

[10] S. Milius, "The mystery of vanishing honeybees is still not definitively solved," Science News, Feb. 13, 2020. Accessed: Nov. 21, 2018. [Online]. Available: https://www.sciencenews.org/article/mystery-vanishing-honeybees-still-not-definitively-solved

[11] A. Tiwari, "A deep learning approach to recognizing bees in video analysis of bee traffic," Ph.D. dissertation, Dept. Comput. Sci., Utah State Univ., Logan, UT, USA, Aug. 2018.

[12] D. D. Reidpath, "Population health. More than the sum of the parts?" *J. Epidemiol. Community Health*, vol. 59, no. 10, pp. 877–880, Oct. 2005, doi: 10.1136/jech.2005.036046.

[13] K. Hon, *An Introduction to Statistics*. Scotts Valley, CA, USA: CreateSpace Independent Publishing Platform, 2014.

[14] L. Chazette, M. Becker, and H. Szczerbicka, "Basic algorithms for bee hive monitoring and laser-based mite control," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Athens, Greece, Dec. 2016, pp. 1–8, doi: 10.1109/SSCI.2016.7850001.

[15] V. Kulyukin, S. Mukherjee, and P. Amlathe, "Toward audio beehive monitoring: Deep learning vs. standard machine learning in classifying beehive audio samples," *Appl. Sci.*, vol. 8, no. 9, p. 1573, Sep. 2018, doi: 10.3390/app8091573.

[16] I. Nolasco and E. Benetos, "To bee or not to bee: Investigating machine learning approaches for beehive sound recognition," in *Proc. Detection Classification Acoust. Scenes Events Workshop (DCASE)*. Woking, U.K.: WWF Living Planet Centre, 2018.

[17] S. Schurischuster, S. Zambanini, and M. Kampel, "Sensor study for monitoring Varroa mites on honey bees (Apis mellifera)," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Cancún, Mexico, 2016.

[18] P. Amlathe, "Standard machine learning techniques in audio beehive monitoring: Classification of audio samples with logistic regression, K-nearest neighbor, random forest and support vector machine," Ph.D. dissertation, Dept. Comput. Sci., Utah State Univ., Logan, UT, USA, May 2018, p. 57.

[19] Australian Government. (2020). *Bees and Bee Pests and Diseases—Department of Agriculture*. Accessed: Nov. 1, 2020. [Online]. Available: https://www.agriculture.gov.au/pests-diseases-weeds/bees

[20] B. Locke, M. Low, and E. Forsgren, "An integrated management strategy to prevent outbreaks and eliminate infection pressure of American foulbrood disease in a commercial beekeeping operation," *Preventive Veterinary Med.*, vol. 167, pp. 48–52, Jun. 2019, doi: 10.1016/j.prevetmed.2019.03.023.

[21] D. Pukhov. (Oct. 2018). *Honey Bee Health Detection With CNN*. Accessed: Aug. 13, 2019. [Online]. Available: https://kaggle.com/dmitrypukhov/honey-bee-health-detection-with-cnn

[22] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5455–5516, Dec. 2020.

[23] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," Mar. 2017, *arXiv:1703.06870*. Accessed: Aug. 15, 2019. [Online]. Available: http://arxiv.org/abs/1703.06870

[24] Utah State University. (2019). *BeePi_Audio_Classification*. Accessed: Aug. 15, 2019. [Online]. Available: https://usu.app.box.com/v/BeePiAudioData

[25] M. Florea, "Automatic detection of honeybees in a hive," M.S. thesis, Dept. Inf. Technol., Univ., Uppsala, Uppsala, Sweden, Sep. 2013.

[26] Y. Sasaki, "The truth of the F-measure," Dept. Comput. Sci., Old Dominion Univ., Norfolk, VA, USA, Tech. Rep., p. 5.

[27] J. Yang. (2018). *The Bee Image Dataset: Annotated Honey Bee Images*. Accessed: Aug. 13, 2019. [Online]. Available: https://kaggle.com/jenny18/honey-bee-annotated-images

[28] R. Willingham, J. Klopchin, and J. Ellis, "Robbing behavior in honey bees," *Behav. Ecol.*, vol. 11, no. 3, pp. 326–333, May 2000, doi: 10.1093/beheco/11.3.326.

[29] W. G. Meikle and N. Holst, "Application of continuous monitoring of honeybee colonies," *Apidologie*, vol. 46, no. 1, pp. 10–22, Jan. 2015, doi: 10.1007/s13592-014-0298-x.

[30] Explore. (May 2019). *Honey Bees Landing Zone Camera—Live Video of Bees*. Accessed: Aug. 14, 2019. [Online]. Available: https://explore.org/livecams/honey-bees/honey-bee-landing-zone-cam

[31] Explore. (May 2019). *Bee Cam—Live Camera Inside of a Honey Bee Hive*. Accessed: Aug. 14, 2019. [Online]. Available: https://explore.org/livecams/honey-bees/honey-bee-hive-cam

[32] M. Filannino and M. Di Bari, "Gold standard vs. silver standard: The case of dependency parsing for Italian," in *Proc. CLiC*, Nov. 2016, p. 288, doi: 10.4000/books.aaccademia.1475.

[33] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, ‘‘The balanced accuracy and its posterior distribution,’’ in *Proc. 20th Int. Conf. Pattern Recognit.*, Istanbul, Turkey, Aug. 2010, pp. 3121–3124, doi: 10.1109/ICPR.2010.764.

[34] G. Tennenholtz, T. Zahavy, and S. Mannor, ‘‘Train on validation: Squeezing the data lemon,’’ Feb. 2018, *arXiv:1802.05846*. Accessed: Aug. 14, 2019. [Online]. Available: http://arxiv.org/abs/1802.05846

[35] C. A. Ramezan, T. A. Warner, and A. E. Maxwell, ‘‘Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification,’’ *Remote Sens.*, vol. 11, no. 2, p. 185, Jan. 2019, doi: 10.3390/rs11020185.

DIOGO BRAGA was born in Oporto, in 1996. He received the Associate degree in informatics engineering and the master’s degree in the context of bee health monitoring systems from the Institute of Engineering–Polytechnic of Porto (ISEP/IPP), in 2017 and 2020, respectively.

He is currently a member of the Interdisciplinary Studies Research Center (ISRC) and has previously conducted work in the field of neurodegenerative diseases.

ANA MADUREIRA (Senior Member, IEEE) was born in Mozambique, in 1969. She received the B.Sc. degree in computer engineering from ISEP, in 1993, the master’s degree in electrical and computers engineering–industrial informatics from FEUP, in 1996, and the Ph.D. degree in production and systems from the University of Minho, Portugal, in 2003.

She is currently a Coordinator Professor at the Institute of Engineering–Polytechnic of Porto (ISEP/IPP) and the Director of the Interdisciplinary Studies Research Center (ISRC). She is the author of more than 100 scientific articles in scientific conference proceedings, journals, and books, in the last few years. She has been an External Member Evaluation Committee of the Agency for Assessment and Accreditation of Higher Education - A3ES for the Scientific Area of Informatics of Polytechnic Higher Education, since 2012. She was the Chair of IEEE Portugal Section, from 2015 to 2017, and the Chair/Vice-Chair of IEEE-CIS Portuguese Chapter. She was the Department of IEEE R8 Educational Activities Sub-Committee, Chair of University, from 2017 to 2018. She was an IEEE R8 Secretary, from 2019 to 2020.

FABIO SCOTTI (Senior Member, IEEE) received the Ph.D. degree in computer engineering from Politecnico di Milano, Milan, Italy, in 2003.

He was an Assistant Professor at the Department of Information Technologies, Università degli Studi di Milano, Italy, from 2002 to 2015. He was an Associate Professor at the Department of Computer Science, Università degli Studi di Milano, Italy, from 2015 to 2020. He has been a Full Professor at the Università degli Studi di Milano, since 2020. Original results have been published in over 130 articles in international journals, proceedings of international conferences, books, book chapters, and patents. His current research interests include biometric systems, machine learning and computational intelligence, signal and image processing, theory and applications of neural networks, three-dimensional reconstruction, industrial applications, intelligent measurement systems, and high-level system design.

Dr. Scotti is an Associate Editor of the IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS and the IEEE OPEN JOURNAL OF SIGNAL PROCESSING. He is serving as a Book Editor (an Area Editor and a Section Less-constrained Biometrics) of the *Encyclopedia of Cryptography, Security, Privacy* (Springer, 3rd Edition). He has been an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY and *Soft Computing* (Springer) and a Guest Coeditor for the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT.

VINCENZO PIURI (Fellow, IEEE) received the Ph.D. degree in computer engineering from Politecnico di Milano, Italy, in 1989.

He has been a Full Professor of computer engineering at the Università degli Studi di Milano, Italy, since 2000. He has also been an Associate Professor at Politecnico di Milano, and a Visiting Professor at The University of Texas at Austin, USA, and a Visiting Researcher at the George Mason University, USA. He is an Honorary Professor at Obuda University, Hungary, the Guangdong University of Petrochemical Technology, China, the Northeastern University, China, Muroran Institute of Technology, Japan, and Amity University, India. Original results have been published in more than 400 articles in international journals, proceedings of international conferences, books, and book chapters. His research interests include artificial intelligence, computational intelligence, intelligent systems, machine learning, pattern analysis and recognition, signal and image processing, biometrics, intelligent measurement systems, industrial applications, digital processing architectures, fault tolerance, dependability, and cloud computing infrastructures.

Dr. Piuri is a Senior Member of INNS. He received the IEEE Instrumentation and Measurement Society Technical Award, in 2002, and the IEEE TAB Hall of Honor, in 2019. He is the President of the IEEE Systems Council, from 2020 to 2021, and he has been the IEEE Vice President for Technical Activities, in 2015, the IEEE Director, the President of the IEEE Computational Intelligence Society, the Vice President for Education of the IEEE Biometrics Council, the Vice President for Publications of the IEEE Instrumentation and Measurement Society and the IEEE Systems Council, and the Vice President for Membership of the IEEE Computational Intelligence Society. He was an Editor-in-Chief of the IEEE SYSTEMS JOURNAL, from 2013 to 2019, and an Associate Editor of the IEEE TRANSACTIONS ON CLOUD COMPUTING. He has been an Associate Editor of IEEE TRANSACTIONS ON COMPUTERS, IEEE TRANSACTIONS ON NEURAL NETWORKS, IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, and IEEE ACCESS. He is also a Distinguished Scientist of ACM.

AJITH ABRAHAM (Senior Member, IEEE) received the M.Sc. degree from Nanyang Technological University, Singapore, in 1998, and the Ph.D. degree in computer science from Monash University, Melbourne, VIC, Australia, in 2001.

He is currently the Director of Machine Intelligence Research Labs (MIR Labs), a Not-for-Profit Scientific Network for Innovation and Research Excellence connecting Industry and Academia. The Network with HQ in Seattle, WA, USA, has currently more than 1 500 scientific members from over 105 countries. As an Investigator/Co-Investigator, he has won research grants worth over 100+ Million U.S. He works in a multi-disciplinary environment and he has authored/coauthored more than 1 400+ research publications out of which there are 100+ books covering various aspects of computer science. One of his books was translated to Japanese and few other articles were translated to Russian and Chinese. He has more than 40 000+ academic citations (H-index of 95 as per Google Scholar). He has given more than 100 plenary lectures and conference tutorials (in 20+ countries).

Dr. Abraham has been the Chair of IEEE Systems Man and Cybernetics Society Technical Committee on Soft Computing (which has over 200+ members), since 2008. From 2011 to 2013, he served as a Distinguished Lecturer of IEEE Computer Society representing Europe. He is an Editor-in-Chief of *Engineering Applications of Artificial Intelligence* (EAAI) and serves/served for the editorial board of over 15 international journals indexed by Thomson ISI.