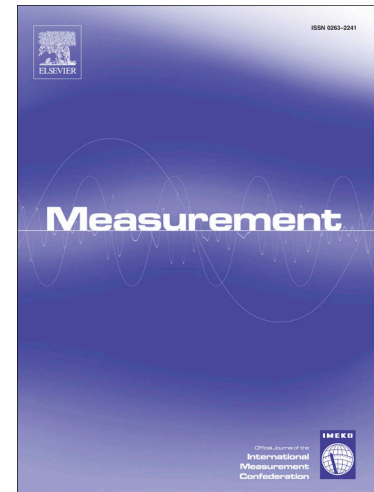# Journal Pre-proofs

A Hybrid Deep Transfer Learning Model with Machine Learning Methods for Face Mask Detection in the Era of the COVID-19 Pandemic

Mohamed Loey, Gunasekaran Manogaran, Mohamed Hamed N. Taha, Nour Eldeen M. Khalifa

Article

A Hybrid Deep Transfer Learning Model with Machine Learning Methods for Face Mask Detection in the Era of the COVID-19 Pandemic

Mohamed Loey [1*], Gunasekaran Manogaran [2,3], Mohamed Hamed N. Taha[4], Nour Eldeen M. Khalifa [4]

[1]   Department of Computer Science, Faculty of Computers Artificial Intelligence, Benha University, Benha 13511, Egypt; email: mloey@fci.bu.edu.eg
[2]   University of California, Davis, USA; email: gmanogaran@ieee.org
[3]   College of Information and Electrical Engineering, Asia University, Taiwan
[4]   Department of Information Technology, Faculty of Computers & Artificial Intelligence, Cairo University, Cairo 12613, Egypt; email: {nourmahmoud, mnasrtaha}@cu.edu.eg

*Correspondence Author: mloey@fci.bu.edu.eg.

Abstract: The coronavirus COVID-19 pandemic is causing a global health crisis. One of the effective protection methods is wearing a face mask in public areas according to the World Health Organization (WHO). In this paper, a hybrid model using deep and classical machine learning for face mask detection will be presented. The proposed model consists of two components. The first component is designed for feature extraction using Resnet50. While the second component is designed for the classification process of face masks using decision trees, Support Vector Machine (SVM), and ensemble algorithm. Three face masked datasets have been selected for investigation. The Three datasets are the Real-World Masked Face Dataset (RMFD), the Simulated Masked Face Dataset (SMFD), and the Labeled Faces in the Wild (LFW). The SVM classifier achieved 99.64 % testing accuracy in RMFD. In SMFD, it achieved 99.49%, while in LFW, it achieved 100% testing accuracy.

Keywords: COVID-19, Masked Face, Deep Transfer Learning, Classical Machine Learning.

1.   Introduction

The trend of wearing face masks in public is rising due to the COVID-19 coronavirus epidemic all over the world. Before Covid-19, People used to wear masks to protect their health from air pollution. While other people are self-conscious about their looks, they hide their emotions from the public by hiding their faces. Scientists proofed that wearing face masks works on impeding COVID-19 transmission [1]. COVID-19 (known as coronavirus) is the latest epidemic virus that hit the human health in the last century [2].   In 2020, the rapid spreading of COVID-19 has forced the World Health Organization to declare COVID-19 as a global pandemic. According to [3], more than five million cases were infected by COVID-19 in less than 6 months across 188 countries. The virus spreads through close contact and in crowded and overcrowded areas.

The coronavirus epidemic has given rise to an extraordinary degree of worldwide scientific cooperation. Artificial Intelligence (AI) based on Machine learning and Deep Learning can help to fight Covid-19 in many ways. Machine learning allows researchers and clinicians evaluate vast quantities of data to forecast the distribution of COVID-19, to serve as an early warning mechanism for potential pandemics, and to classify vulnerable populations. The provision of healthcare needs funding for emerging technology such as artificial intelligence, IoT, big data and machine learning to tackle and predict new diseases. In order to better understand

infection rates and to trace and quickly detect infections, the AI 's power is being exploited to address the Covid-19 pandemic [4] such as the detection of COVID-19 in medical chest X-rays [5].

Policymakers are facing a lot of challenges and risks in facing the spreading and transmission of COVID-19 [6]. People are forced by laws to wear face masks in public in many countries. These rules and laws were developed as an action to the exponential growth in cases and deaths in many areas. However, the process of monitoring large groups of people is becoming more difficult. The monitoring process involves the detection of anyone who is not wearing a face mask. In France, to guarantee that riders wear face masks, new AI software tools are integrated in the Paris Metro system's surveillance cameras [7]. The French startup DatakaLab [8], which developed the software, reports that the goal is not to recognize or arrest people who do not wear masks but to produce anonymous statistical data that can help the authorities predict potential outbreaks of COVID-19.

In this paper, we introduce a mask face detection model that is based on deep transfer learning and classical machine learning classifiers. The proposed model can be integrated with surveillance cameras to impede the COVID-19 transmission by allowing the detection of people who are not wearing face masks. The model is integration between deep transfer learning and classical machine learning algorithms. We have used deep transfer leering for feature extractions and combined it with three classical machine learning algorithms. We introduced a comparison between them to find the most suitable algorithm that achieved the highest accuracy and consumed the least time in the process of training and detection.

The novelty of this research is using a proposed feature extraction model have an end-to-end structure without traditional techniques with three classifiers machine learning algorithms for mask face detection. The organization for the rest of the paper is as follows. Section 2 reviews previous related works. Section 3 describes the characteristics of the dataset. Section 4 illustrates the proposed model in detail. Section 5 reports and analyses the experimental results, and Section 6 presents the conclusions and possibilities of future work.

2.   Related works

Generally, most of the publication focus is on face construction and identity recognition when wearing face masks. In this research our focus is on recognizing the people who are not wearing face masks to help in decreasing the transmission and spreading of the COVID-19. Researchers and scientists have proved that wearing face masks help in minimizing the spreading rate of COVID-19. In [9], the authors developed a new facemask-wearing condition identification method. They were able to classify three categories of facemask-wearing conditions. The categories are correct facemask-wearing, incorrect facemask-wearing, and no facemask-wearing. The proposed mothed has achieved 98.70% accuracy in the face detection phase. Sabbir et al [10], have applied the Principal Component Analysis (PCA) on masked and unmasked face recognition to recognize the person. They found that the accuracy of face resonation using the PCA is extremity affected by wearing masks. The recognition accuracy drops to less than 70% when the recognized face is masked. Also, PCA was used in [11]. The authors proposed a method that is used for removing glasses from a human frontal facial image. The removed part was reconstructed using recursive error compensation using PCA reconstruction.

In [12], the authors used the YOLOv3 algorithm for face detection. YOLOv3 uses Darknet-53 as the backbone. The proposed method achieved 93.9% accuracy. It was trained on CelebA and WIDER FACE dataset including more than 600,000 images. The testing was the FDDB dataset. Nizam et al [13] proposed a novel GAN-based

network that can automatically remove masks covering the face area and regenerate the image by building the missing hole. The output of the proposed model is a complete face image that looks natural and realistic.

In [14], the authors presented a system for detecting the presence or absence of a compulsory medical mask in the operating room. The overall objective is to minimize the false positive face detections as possible without missing mask detections in order to trigger alarms only for medical staff who do not wear a surgical mask. The proposed system archived 95% accuracy.

Muhammad et al [15] presented an interactive method called MRGAN. The method depends on getting the microphone area from the user and using the Generative Adversarial Network to rebuild this area. Shaik et al [16] used deep learning real-time face emotion classification and recognition. They used VGG-16 to classify seven facial expressions. The proposed model was trained on the KDEF dataset and achieved 88% accuracy.

3. Datasets Characteristics

This research conducted its experiments on three original datasets. The first dataset is Real-World Masked Face Dataset (RMFD) [17]. The author of RMFD created one of the biggest masked face datasets used in this research. The RMFD dataset consists of 5000 masked faces and 90,000 unmasked faces. Figure 1 illustrates samples of faces with and without masks. In this research, 5000 images for faces with masks and without masks have been used with a total of 10000 images to balance the dataset. The RMFD dataset used for the training, validation, and testing phases.
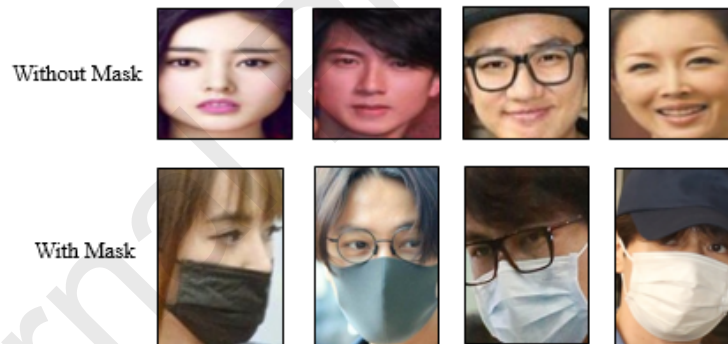

Fig. 1. RMFD dataset images samples

The second dataset is a Simulated Masked Face Dataset (SMFD) [18]. The SMFD dataset consists of 1570 images, 785 for simulated masked faces, 785 for unmasked faces. Examples for images of the SMFD are presented in Figure 2. The SMFD dataset used for the training, validation, and testing phases.

Fig. 2. SMFD dataset images samples

The Third dataset used in this research is the Labeled Faces in the Wild (LFW) [19]. It is a simulated masked face dataset that contains 13000 masked faces for celebrities around the round. Figure 3 illustrates samples of LFW images. The LFW dataset used for the testing phase only as a benchmark testing dataset which the proposed model never trained on it.

Fig. 3. LFW dataset images samples

4. The Proposed Model

The introduced model includes two main components, the first component is deep transferring learning (ResNet50) as feature extractor and the second component is a classical machine learning like decision trees, SVM, and ensemble. According to [20] [21], ResNet-50 has achieved better results when it is used as a feature extractor. Figure 4 illustrates the proposed classical transfer learning model. Mainly, the ResNet50 used for the feature extraction phase while the traditional machine learning model used in the training, validation, and testing phase.
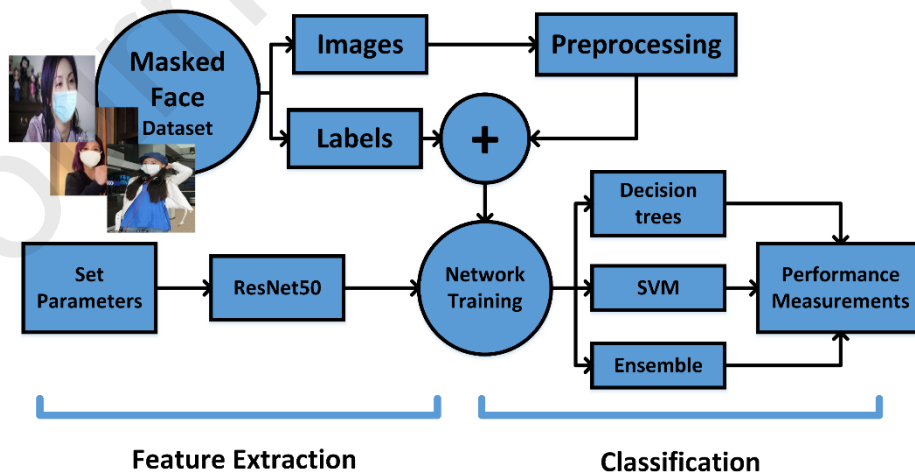
Fig 4. The proposed deep transfer learning model.

A residual neural network (ResNet) is a kind of deep transfer learning based on residual learning [22]. All types of ResNet-101, ResNet-50, and ResNet-18 are versions of ResNet to get rid of the problem of vanishing

gradients that have their specific residual block. ResNet-50 with 50-layers are deep, start with a convolution layer, and end with a fully-connected layer, and in between followed by 16 residual bottleneck blocks each block has three layers of convolution layer as shown in Figure 5.

In classification, the last layer in ResNet-50 was removed and replaced with three traditional machine learning classifiers (Support vector machine (SVM), decision tree, and ensemble) to improve our model performance. The main contribution of this research is to construct SVM, decision trees, and ensemble that do not overfit the training process.
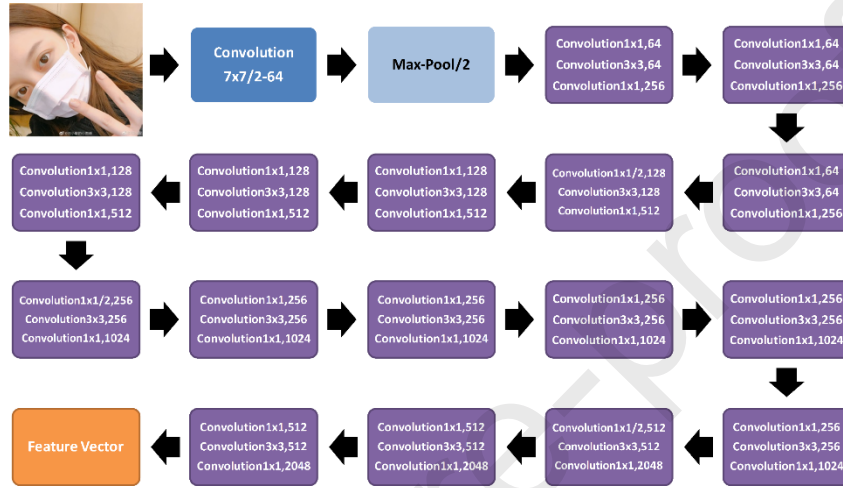

Fig 5. Proposed ResNet-50 as the feature extractor

4.1   Support Vector Machine

One of the most popular and spectacular supervised learning techniques with related learning algorithms for treatment classification and regression tasks in patterns is SVM. SVM is a classification machine learning algorithm based on hinge function as shown in Equation 1, where z is a label from 0 to 1, $w.I - b$ is the output, w and b are coefficients of linear classification, and I is an input vector. The loss function to be minimized can be implemented in Equation 2 [23], [24].

$$h_j = \max\left(0, 1 - z_j(w.I_j - b)\right) \tag{1}$$

$$loss = \frac{1}{n}\sum_{i=1}^{n}\max\left(0, h_i\right) \tag{2}$$

4.2   Decision Tree

The decision tree is the classification model of computation based on entropy function and information gain. Entropy computes the amount of uncertainty in data as shown in Equation 3. Where D is current data, and q is

a binary label from 0 to 1, and p(x) is the proportion of q label. To measure the difference of entropy from data, we calculate information gain (I) as illustrated in equation 4. Where v is a subset of data [25], [26].

$$E(D) = \sum_{i=1}^{m} -p(q_i) \cdot \log(p(q_i)) \tag{3}$$

$$I = E(D) - \sum_{v \in D} p(v)E(v) \tag{4}$$

4.3   Ensemble Methods

Ensemble methods are algorithms of machine learning that create a collection of classifiers. An ensemble of classifiers is a collection of classifiers whose individual decisions (usually by weighted or unweighted voting) are merged in one way or another to identify new instances [27]. The used Ensemble methods are K-Nearest Neighbors Algorithm (k-NN) [28], Linear Regression [29] and Logistic Regression [30]. The steps of the ensemble method are: 1) generate M classifiers 2) Train each classifier alone 3) merge the M classifiers and average their output. We improve our ensemble by using complex weight (α) to get better results as illustrated in equation 5 [31].

$$\bar{z} = \sum_{i=1}^{M} \alpha_i z_i \tag{5}$$

5.   Experimental Results

All the experimental trials have been conducted on a computer sever equipped by an Intel Xeon processor (2 GHz), 96 GB of RAM. The MATLAB software package was selected in this research for the development and implementation of the different experimental trails. The experiments trails include the following specifications and setup:

- Three classifiers (Decision trees, SVM, and Ensemble).
- Three datasets:
  - A dataset of RMFD with real face masks for (training, and testing phases), will be referred to DS1.
  - A dataset of SMFD with fake face masks for (training, and testing phases), will be referred to DS2.
  - A combined dataset from DS1, and DS2 for (training, and testing phases), will be referred to DS3.
  - A dataset of LFW with simulated face masks for (testing), will be referred to DS4.
- Datasets for the (training, and testing) are split up to (70% for training, 10% for validation, 20% for testing phase)

To evaluate the performance of the different classifiers, performance matrices are needed to be investigated through this research. The most common performance measures to be calculated are Accuracy, Precision, Recall, and F1 Score [32], and they are presented from Equation (6) to Equation (9).

$$\text{Accuracy} = \frac{TP + TN}{(TP + FP) + (TN + FN)} \tag{6}$$

$$\text{Precision} = \frac{TP}{(TP + FP)} \tag{7}$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \tag{8}$$

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{(\text{Precision} + \text{Recall})} \tag{9}$$

Where TP is the count of True Positive samples, TN is the count of True Negative samples, FP is the count of False Positive samples, and FN is the count of False Negative samples from a confusion matrix. The authors of this research build their deep transfer learning based on [33]–[38] to improve image classification accuracy, but the classification accuracy wasn't acceptable. The experimental results will be presented in five subsections, and the first subsection will discuss the achieved results for the decision trees classifier while the second subsection will introduce the results for the SVM classifier. Subsection number three will present the obtained results for the ensemble classifier. Subsection four will illustrate the confusion matrices for the different classifiers. Finally, the fifth subsection will illustrate a comparative results analysis with related works according to the testing accuracy.

5.1 Validation, Testing Accuracy, and Performance Metrics for Decision Trees Classifier

As mentioned earlier in the experimental setup, three datasets (DS1, DS2, and DS3) will be experimented on for training, validation and testing. The DS4 will be used for testing only. Figure 6 illustrates the achieved results for the decision trees classifier in the validation phase for the different datasets.
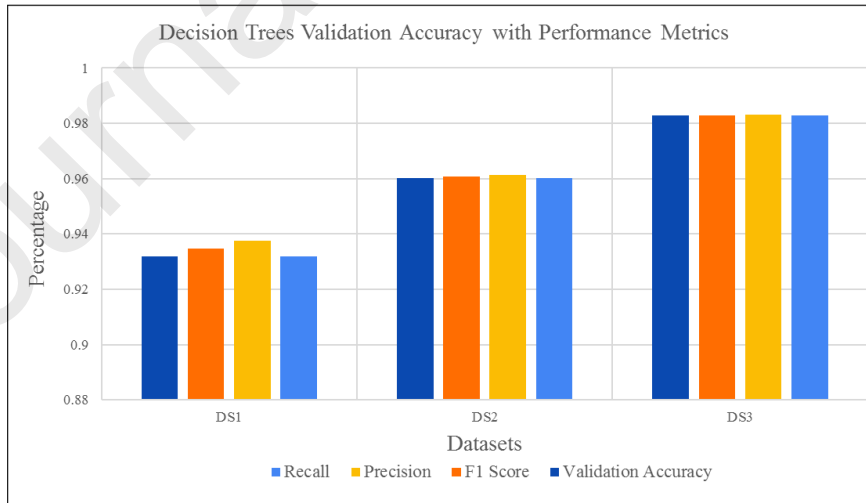

Fig. 6. Decision trees classifier validation accuracy with performance metrics for the different datasets.

For the DS1, the decision trees classifier achieved a validation accuracy with performance metrics ranging from 92% to 94%. The DS1 contains real masks with different poses for face and different masks types. Moreover, the decision trees classifier in DS2 achieved a percentage of 96% for the validation accuracy with performance metrics. The DS2 contains fake masks over real faces images. In DS3, the decision trees classifier achieved a

percentage of 98% for the validation accuracy with performance metrics. DS3 is a combined dataset from DS1 and DS2. DS3 is a large dataset in terms of the number of images which help in achieving better accuracies, and more data means better accuracies in machine learning [39]. Although the time is relative from machine to another machine, it is a good indicator to measure the performance of the classifier [40]. Figure 7 illustrates the time consumed by the decision trees classifier for the training process for the different datasets.
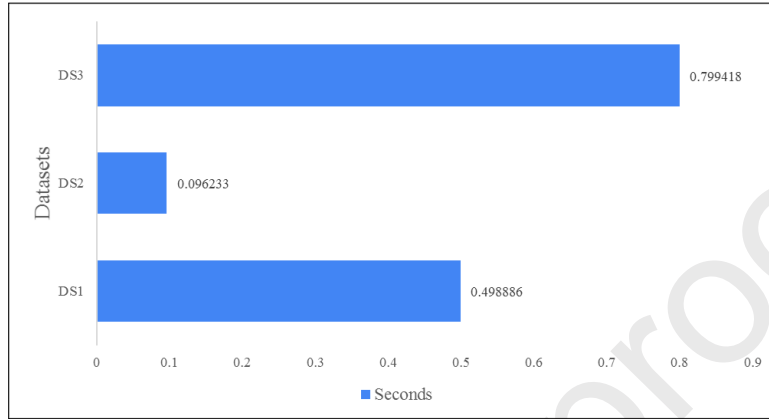

Fig. 7. Consumed time for the training process for the decision trees classifier for different datasets.

The consumed time is relative to the dataset size and machine capabilities. The dataset's number of images is illustrated in the dataset characteristics section. To evaluate the performance of the decision trees classifier, different testing strategies has been tested through this research and they are summarized as follow:

- Training over DS1, and testing over DS1, DS2, DS3, and DS4.
- Training over DS2, and testing over DS1, DS2, DS3, and DS4.
- Training over DS3, and testing over DS1, DS2, DS3, and DS4.

Figure 8 illustrates the achieved percentage for the testing accuracy and performance metrics for the different testing strategies for the decision trees classifier.
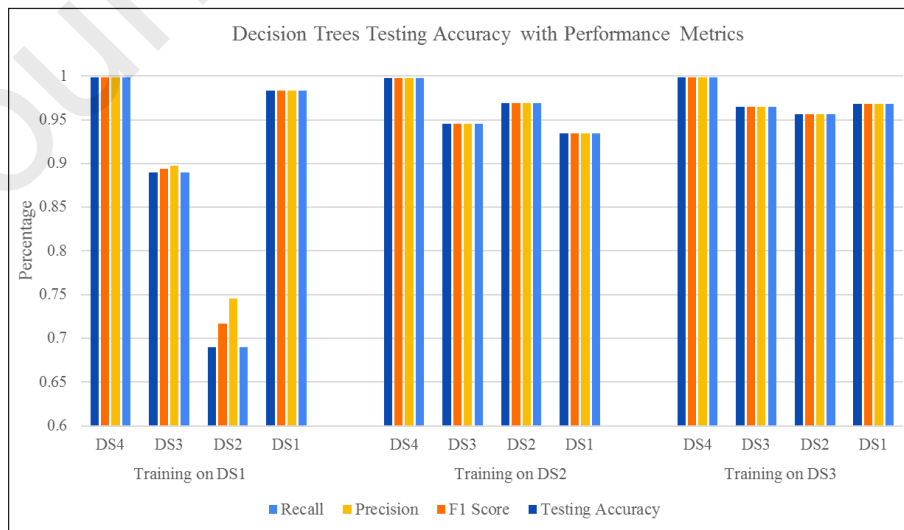

Fig. 8. Testing accuracy with performance metrics for the decision trees classifier with different testing strategies.

Figure 8 shows exciting results, and they are 1) on the training over DS1, the decision trees classifier wasn't able to achieve a good classification accuracy 68% in DS2, as DS2 contains a lot of variation of fake masks. That's also will reflect in the DS3 which is a combined dataset from DS1, and DS2. 2) on the training over DS2, the decision trees classifier was able to achieve 93% for DS1, which contains real masks. 3) on the training over DS3, the decision trees classifier achieved the highest accuracy with performance metrics in all datasets. All the achieved results are above 95%. 4) on the DS4 which is used only for testing and never been trained on it, the decision trees classifier achieved a competitive accuracy with 99% whatever the training is performed over DS1, DS2 or DS3.

From this subsection, we conclude that the decision trees classifier achieved the highest accuracy possible when the training is performed over the DS3. The highest testing accuracy for DS1, DS2, DS3, and DS4 was 96.78%, 95.64%, 96.5%, and 99.89% respectively.

5.2 Validation, Testing Accuracy, and Performance Metrics for Decision SVM Classifier

The same experimental trials which were conducted using decision trees classifiers will be performed on the SVM classifier. Figure 9 presents the validation accuracy and performance metrics for the SVM classifier for the different datasets.
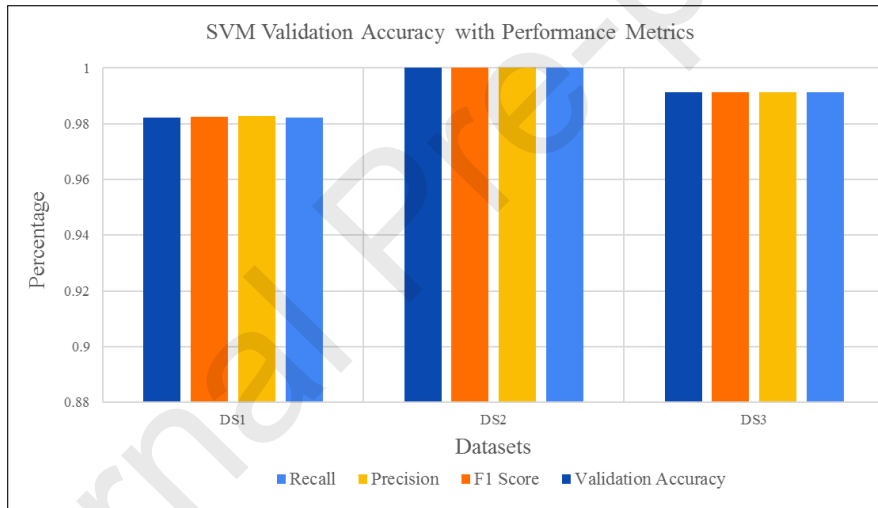

Fig. 9. SVM classifier validation accuracy with performance metrics for the different datasets.

Figure 9 shows that the SVM classifier achieved a higher validation accuracy for all datasets than the decision trees classifier. In DS1, SVM achieved 98% while decision trees achieved 93% in the validation accuracy. In DS2, the SVM classifier achieved 100% while decision trees achieved 96%. In DS3, SVM achieved 99% while decision trees achieved 98%. The SVM classifier surpasses the validation accuracy along with performance metrics than the decision trees classifier. One more notable remark, training over the DS2 achieved the highest validation accuracy possible with 100% accuracy while in the decision trees classifier the highest validation accuracy was 98% in DS3.

The consumed time also is an essential factor in evaluating the performance of the classifier, and Figure 10 illustrated the wasted time for the SVM classifier for the different datasets.
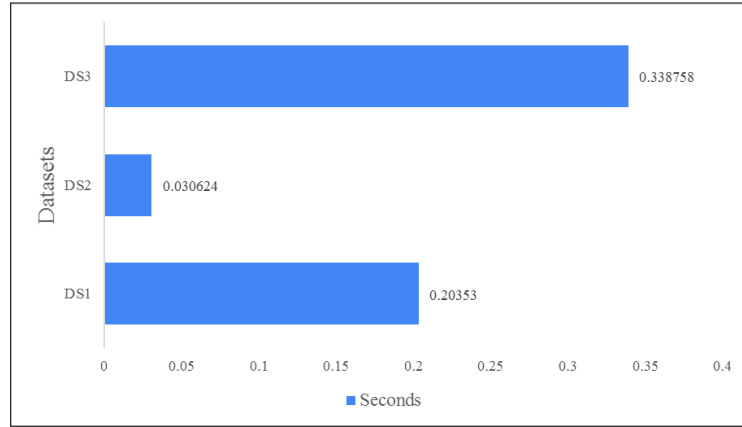
Fig. 10. Consumed time for the training process for the SVM classifier for the different datasets.

Using the SVM classifier, the more the data exists, the more time the classifier will consume time, DS3 contains the largest number of images among the introduced datasets, so the DS3 consumes more time in the training process. A notable remark that the consumed time for the SVM classifier is less than the decision trees classifier for all the datasets. In the DS1, The SVM classifier consumed less time than decision trees classifier by 0.29 seconds (improvement by 59%). While in the DS2, The SVM classifier consumed less time than the decision trees classifier by 0.06 seconds (improvement by 68%). In the DS3, The SVM classifier consumed less time than the decision trees classifier by 0.06 seconds (improvement by 57%).

Figure 11 illustrates the achieved percentage for the testing accuracy and performance metrics of the SVM classifier for the different testing strategies which were introduced in the decision trees classifier section.
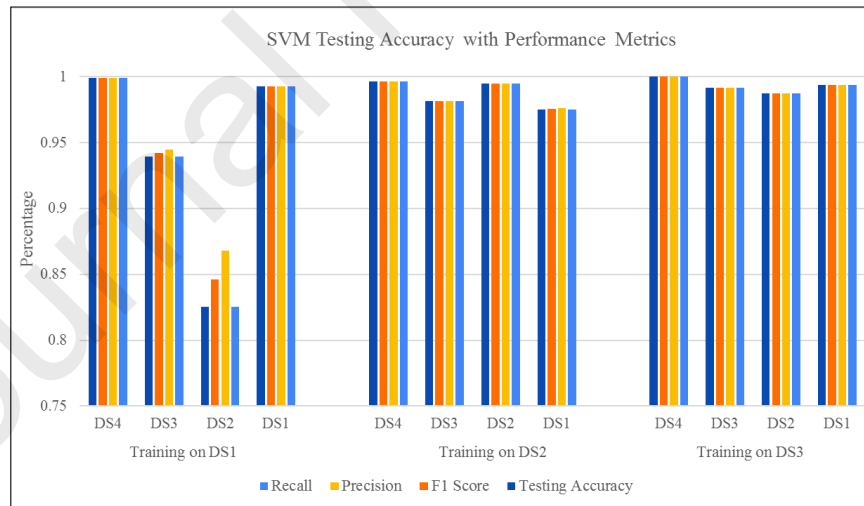

Fig. 11. Testing accuracy with performance metrics for the SVM classifier with different testing strategies.

Figure 11 shows acceptable results and they are 1) the behavior of the SVM classifier is similar to the decision trees classifier but the SVM classifier achieves a higher testing accuracy. On the training over DS1, the SVM classifier, achieved 82% over DS2 in the testing accuracy while decision trees classifier achieved 68%. On the training over DS2, the SVM classifier, achieves higher accuracies over 97% for all datasets, while the decision trees classifier achieved accuracies over 93%. The same pattern happens for the training over DS3, the SVM classifier, achieved higher accuracies over 98% for all datasets, while the decision trees classifier achieved

accuracies over 95%. 2) on the DS4 which is used only for testing and never been trained on, the SVM classifier achieved a higher accuracy with 99% whatever the training is performed over DS1, DS2 or DS3.

From this subsection, we conclude according to the achieved results that the SVM classifier is better than decision trees classifier in terms of validation, testing accuracy, performance metrics, and consumed time. The highest testing accuracy for DS1 was achieved by training over DS3 with 99.4%. For DS2, the highest accuracy was achieved by training over DS2 with 99.49%. In DS3, the highest accuracy was achieved by training over DS3 with 99.19%, and for DS4, the highest testing accuracy was achieved by training over DS3 with 100%.

5.3   Validation, Testing Accuracy, and Performance Metrics for The Ensemble Classifier

The same experimental trials which were conducted on decision trees and SVM classifier will be performed on the ensemble classifier. Figure 12 presents the validation accuracy and performance metrics for the ensemble classifier for the different datasets.
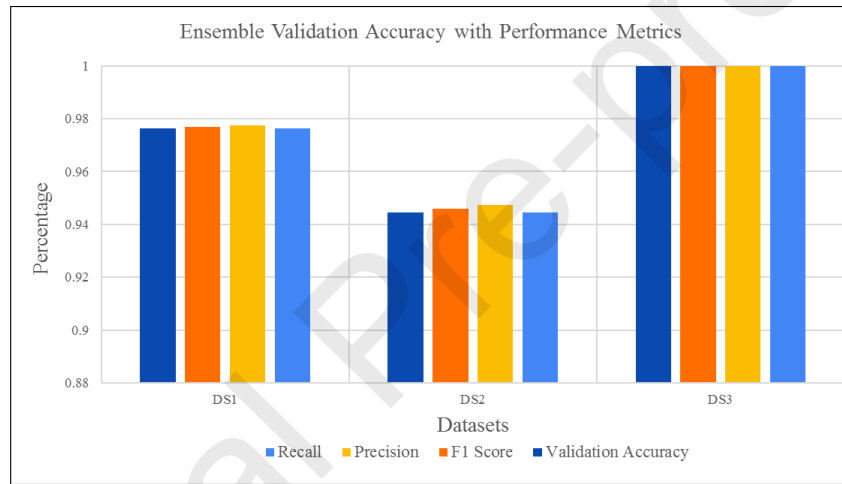

Fig. 12. Ensemble classifier validation accuracy with performance metrics for the different datasets

Figure 12 illustrated that the ensemble classifier achieves the highest accuracy when there are more data exist for the training. The ensemble classifier achieved 100% in the testing accuracy for DS3. It outperformed the decision trees and the SVM classifier. While in DS2, The SVM achieved the highest validation accuracy, with 100%, the ensemble classifier achieved 94%. In DS1, The SVM achieved the highest validation accuracy, with 98% percent, while the ensemble classifier achieved 97%.

Figure 13 presents the consumed time for the training process for the ensemble classifier. Figure 7 shows clearly that the ensemble wasted more time than the decision trees and the SVM classifier. The consumed time for the training of DS3 is 29.2 seconds while for the SVM, and the decision trees classifier were 0.33, and 0.79 seconds. From the achieved results, we can conclude that the ensemble classifier performance according to consumed time is not competitive at all. This is due to the nature of the ensemble classifier to try all possible classifiers that achieve the highest accuracy, which by definition takes a long time if it is compared to the other classifiers.
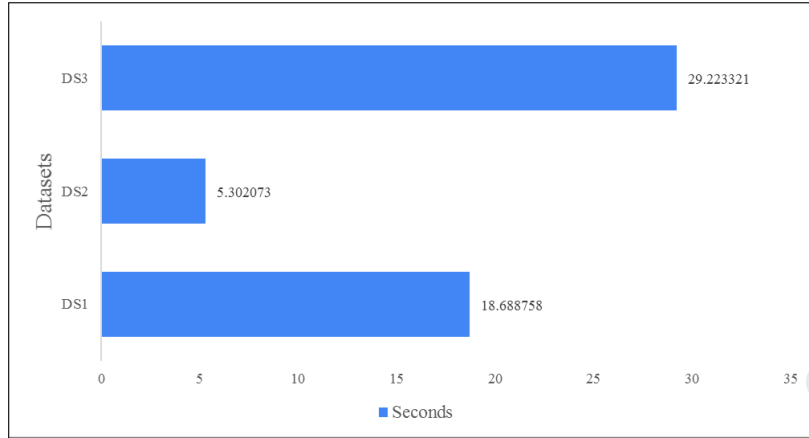
Fig. 13. Consumed time for the training process for the ensemble classifier for the different datasets.

Figure 14 illustrates the achieved percentage for the testing accuracy and performance metrics of the ensemble classifier for the different testing strategies which were introduced in the decision trees classifier section.
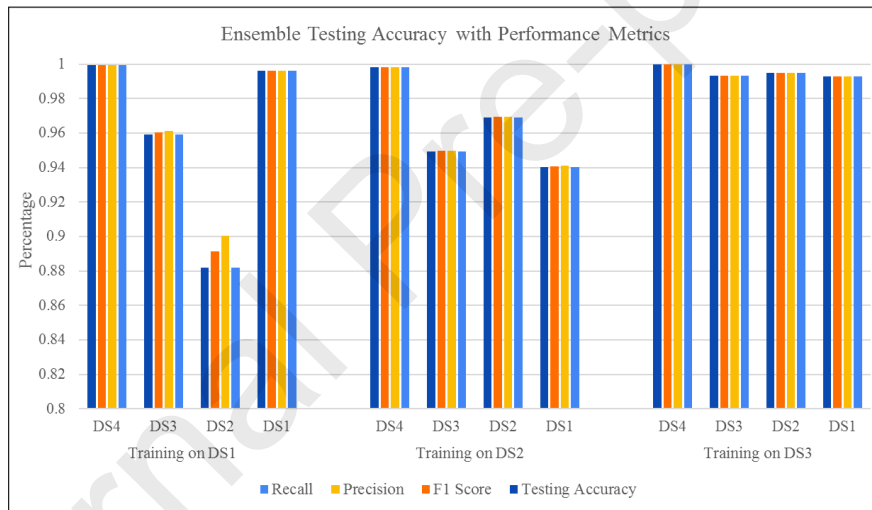

Fig. 14. Testing accuracy with performance metrics for the ensemble classifier with different testing strategies.

Relevant results appear in Figure 14, and they are 1) on the training over DS1, the ensemble classifier outperformed the SVM classifier, it achieved 99.64%, 88.21%, 95.93, and 99.95% for DS1, DS2, DS3, and DS4 respectively. While the SVM classifier achieved 99.28%, 82.56%, 93.97%, and 99.92% for DS1, DS2, DS3 and DS4 accordingly. 2) on the training over DS2, the SVM classifier achieved higher testing accuracy than the ensemble classifier. 3) on the training over DS3, the ensemble classifier outperformed the SVM classifier, it achieved 99.28%, 99.49%, 99.35%, and 100% for DS1, DS2, DS3, and DS4 respectively. While the SVM classifier achieved 99.27%, 98.72%, 99.19%, and 100% for DS1, DS2, DS3 and DS4 respectively. 4) on the DS4 which is used only for testing and never been trained on, the ensemble classifier achieved a higher accuracy over 99% whatever the training is performed over DS1, DS2 or DS3. The ensemble classifier outperformed the SVM classifier in terms of testing accuracy in all training strategies.

From this subsection, we conclude according to the achieved results that the ensemble classifier is better than decision trees, and SVM classifier in terms of validation, testing accuracy, performance metrics when the

training is over DS1, and DS3. While, when the training is over DS2, the SVM classifier outperforms the other classifiers. Moreover, in terms of the consumed time for the training, the SVM classifier is the least consumption of time.

5.4   Confusion Matrix and Class Accuracy for The SVM and Ensemble classifier

Confusion matrices are another useful insight into the performance of the classifiers. Training over combined datasets (DS3) is the most appropriate choice to achieve the highest accuracy possible for the different classifiers. As for decision trees, the confusion matrices are not included in this section as it reached the least testing accuracy. Figure 15 illustrates the confusion matrices for the SVM classifier in the testing phase for DS1, DS2, and DS3 when the training is over DS3.

Fig. 15. Confusion matrix of testing accuracy for (a) DS1, (b) DS2, and (c) DS3 for the SVM classifier over the training of DS3

The achieved testing accuracy for DS1 is 99.4%, for DS2 is 98.7%, and DS3 is 99.2%. Figure 16 illustrates the confusion matrices for the ensemble classifier in the testing phase for DS1, DS2, and DS3 when the training is over DS3.
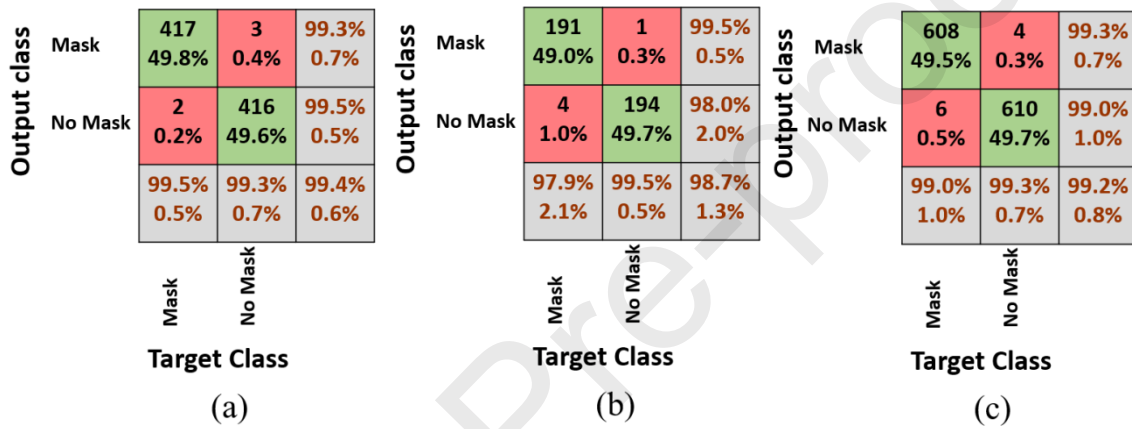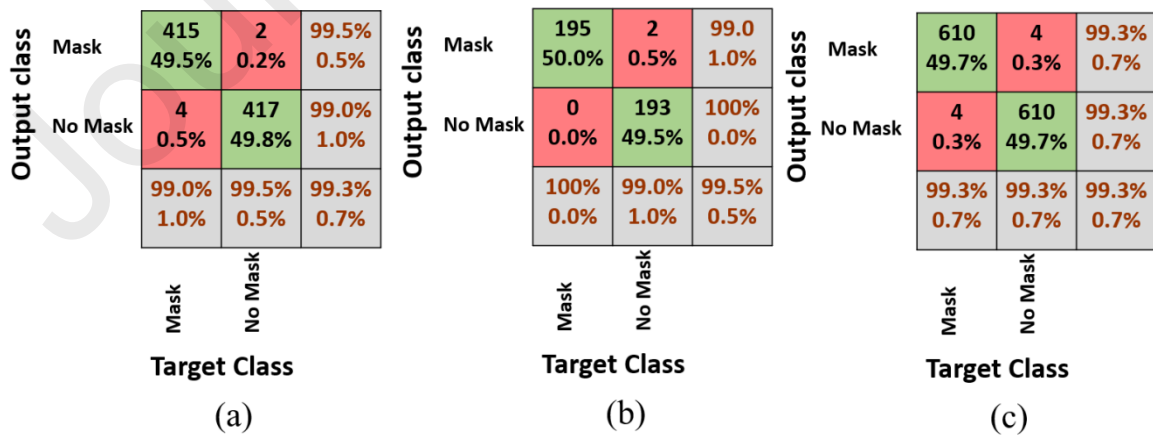

Fig. 16. Confusion matrix of the testing accuracy for (a) DS1, (b) DS2, and (c) DS3 for the ensemble classifier over the training of DS3

The achieved testing accuracy for DS1 is 99.3%, for DS2 is 99.5%, and DS3 is 99.3%. The confusion matrix for DS4 for both SVM and ensemble classifier is the same with 100% testing accuracy. The research decision, according to achieved results which were very close to each other between the SVM, and the ensemble classifier, is to select the SVM classifier for the following reasons:

- All testing accuracy results for training over DS3 is very close only, 0.01 % difference for DS3 between the SVM, and the ensemble classifier.
- In the testing accuracy for DS4, over the training of DS3, the SVM classifier achieved 100% the same result as for the ensemble classifier.
- The SVM classifier consumes less time in training as a performance indicator.

5.5 Comparison with related works

The work presented in [17] used the same datasets, which include the real masked dataset RMFD (DS1) and the fake masked dataset LFW (DS4). The authors of [17] achieved a testing accuracy ranging from 50% to 95%. In the presented work, the testing accuracy for DS1 ranging from 93.44% using the decision tree classifier and 99.64% using the ensemble classifier. For DS4, the testing accuracy ranging from 99.76% using the decision tree classifier and 100% using the SVM classifier.

For the fake masked dataset SMFD (DS2), there is no reported accuracy according to the author of the dataset (https://github.com/prajnasb/observations). In this work, we report an accuracy ranging from 94.54% using the decision tree classifier, and 99.49% using the SVM classifier.

For the combined masked dataset (DS3), there is no reported accuracy according to related works as we present it through this work, we report an accuracy ranging from 96.50% using the decision tree classifier, and 99.35% using the SVM classifier. As a future study, we plan to approach the masked face from a neutrosophic environment with deep transfer learning models.

6. Conclusion and Future Works

The coronavirus COVID-19 pandemic is causing a global health crisis. Governments all over the world are struggling to stand against this type of virus. The protection from infection caused by COVID-19 is a necessary countermeasure, according to the World Health Organization (WHO). In this paper, a hybrid model using deep and classical machine learning for face mask detection was presented. The proposed model consisted of two parts. The first part was for the feature extraction using Resnet50. Resnet50 is one of the popular models in deep transfer learning. While the second part was for the detection process of face masks using classical machine learning algorithms. The Support Vector Machine (SVM), decision trees, and ensemble algorithms were selected as traditional machine learning for investigation.

Three datasets had experimented on, and different training and testing strategies had adopted through this research. The plans include training on a specific dataset while testing over other datasets to prove the efficiency of the proposed model. The presented works concluded that The SVM classifier achieved the highest accuracy possible with the least time consumed in the training process. The SVM classifier in RMFD achieved 99.64 % testing accuracy. In SMFD, it gained 99.49%, while in LFW, it reached 100% testing accuracy. A comparative result had carried out with related works. The proposed model super passed the associated works in terms of testing accuracy. The major drawback is not tray most of classical machine learning methods to get lowest consume time and highest accuracy. One of the possible future tasks is to use deeper transfer learning models

for feature extraction and use the neutrosophic domain as it shows promising potential in the classification and detection problems.

Conflict of interest: The author declares no conflict of interest

References

[1] S. Feng, C. Shen, N. Xia, W. Song, M. Fan, and B. J. Cowling, "Rational use of face masks in the COVID-19 pandemic," *The Lancet Respiratory Medicine*, vol. 8, no. 5, pp. 434–436, May 2020, doi: 10.1016/S2213-2600(20)30134-X.

[2] X. Liu and S. Zhang, "COVID-19: Face masks and human-to-human transmission," *Influenza and Other Respiratory Viruses*, vol. n/a, no. n/a, doi: 10.1111/irv.12740.

[3] "WHO Coronavirus Disease (COVID-19) Dashboard." https://covid19.who.int/ (accessed May 21, 2020).

[4] D. S. W. Ting, L. Carin, V. Dzau, and T. Y. Wong, "Digital technology and COVID-19," *Nature Medicine*, vol. 26, no. 4, Art. no. 4, Apr. 2020, doi: 10.1038/s41591-020-0824-5.

[5] M. Loey, F. Smarandache, and N. E. M Khalifa, "Within the Lack of Chest COVID-19 X-ray Dataset: A Novel Detection Model Based on GAN and Deep Transfer Learning," *Symmetry*, vol. 12, no. 4, p. 651, 2020.

[6] D. M. Altmann, D. C. Douek, and R. J. Boyton, "What policy makers need to know about COVID-19 protective immunity," *The Lancet*, vol. 395, no. 10236, pp. 1527–1529, May 2020, doi: 10.1016/S0140-6736(20)30985-5.

[7] "Paris Tests Face-Mask Recognition Software on Metro Riders," *Bloomberg.com*, May 07, 2020.

[8] "Datakalab | Analyse de l'image par ordinateur." https://www.datakalab.com/ (accessed Jun. 29, 2020).

[9] B. QIN and D. LI, "Identifying Facemask-wearing Condition Using Image Super-Resolution with Classification Network to Prevent COVID-19," May 2020, doi: 10.21203/rs.3.rs-28668/v1.

[10] Md. S. Ejaz, Md. R. Islam, M. Sifatullah, and A. Sarker, "Implementation of Principal Component Analysis on Masked and Non-masked Face Recognition," in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, May 2019, pp. 1–5, doi: 10.1109/ICASERT.2019.8934543.

[11] Jeong-Seon Park, You Hwa Oh, Sang Chul Ahn, and Seong-Whan Lee, "Glasses removal from facial image using recursive error compensation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 805–811, May 2005, doi: 10.1109/TPAMI.2005.103.

[12] C. Li, R. Wang, J. Li, and L. Fei, "Face Detection Based on YOLOv3," in *Recent Trends in Intelligent Computing, Communication and Devices*, Singapore, 2020, pp. 277–284, doi: 10.1007/978-981-13-9406-5_34.

[13] N. Ud Din, K. Javed, S. Bae, and J. Yi, "A Novel GAN-Based Network for Unmasking of Masked Face," *IEEE Access*, vol. 8, pp. 44276–44287, 2020, doi: 10.1109/ACCESS.2020.2977386.

[14] A. Nieto-Rodríguez, M. Mucientes, and V. M. Brea, "System for Medical Mask Detection in the Operating Room Through Facial Attributes," in *Pattern Recognition and Image Analysis*, Cham, 2015, pp. 138–145, doi: 10.1007/978-3-319-19390-8_16.

[15] M. K. J. Khan, N. Ud Din, S. Bae, and J. Yi, "Interactive Removal of Microphone Object in Facial Images," *Electronics*, vol. 8, no. 10, Art. no. 10, Oct. 2019, doi: 10.3390/electronics8101115.

[16] S. A. Hussain and A. S. A. A. Balushi, "A real time face emotion classification and recognition using deep learning model," *J. Phys.: Conf. Ser.*, vol. 1432, p. 012087, Jan. 2020, doi: 10.1088/1742-6596/1432/1/012087.

[17] Z. Wang *et al.*, "Masked face recognition dataset and application," *arXiv preprint arXiv:2003.09093*, 2020.

[18] prajnasb, "observations," *observations*. https://github.com/prajnasb/observations (accessed May 21, 2020).

[19] E. Learned-Miller, G. B. Huang, A. RoyChowdhury, H. Li, and G. Hua, "Labeled Faces in the Wild: A Survey," in *Advances in Face Detection and Facial Image Analysis*, M. Kawulok, M. E. Celebi, and B. Smolka, Eds. Cham: Springer International Publishing, 2016, pp. 189–248.

[20] P. Khojasteh *et al.*, "Exudate detection in fundus images using deeply-learnable features," *Computers in Biology and Medicine*, vol. 104, pp. 62–69, Jan. 2019, doi: 10.1016/j.compbiomed.2018.10.031.

[21] L. Wen, X. Li, and L. Gao, "A transfer convolutional neural network for fault diagnosis based on ResNet-50," *Neural Comput & Applic*, vol. 32, no. 10, pp. 6111–6124, May 2020, doi: 10.1007/s00521-019-04097-w.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[23] A. Çayir, I. Yenidoğan, and H. Dağ, "Feature Extraction Based on Deep Learning for Some Traditional Machine Learning Methods," in *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, Sep. 2018, pp. 494–497, doi: 10.1109/UBMK.2018.8566383.

[24] M. Jogin, Mohana, M. S. Madhulika, G. D. Divya, R. K. Meghana, and S. Apoorva, "Feature Extraction using Convolution Neural Networks (CNN) and Deep Learning," in *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)*, May 2018, pp. 2319–2323, doi: 10.1109/RTEICT42901.2018.9012507.

[25] A. Navada, A. N. Ansari, S. Patil, and B. A. Sonkamble, "Overview of use of decision tree algorithms in machine learning," in *2011 IEEE Control and System Graduate Research Colloquium*, Jun. 2011, pp. 37–42, doi: 10.1109/ICSGRC.2011.5991826.

[26] P.-L. Tu and J.-Y. Chung, "A new decision-tree classification algorithm for machine learning," in *Proceedings Fourth International Conference on Tools with Artificial Intelligence TAI '92*, Nov. 1992, pp. 370–377, doi: 10.1109/TAI.1992.246431.

[27] R. Polikar, "Ensemble Learning," in *Ensemble Machine Learning: Methods and Applications*, C. Zhang and Y. Ma, Eds. Boston, MA: Springer US, 2012, pp. 1–34.

[28] E. Mangalova and E. Agafonov, "Wind power forecasting using the k-nearest neighbors algorithm," *International Journal of Forecasting*, vol. 30, no. 2, pp. 402–406, Apr. 2014, doi: 10.1016/j.ijforecast.2013.07.008.

[29] I. Naseem, R. Togneri, and M. Bennamoun, "Linear Regression for Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 2106–2112, Nov. 2010, doi: 10.1109/TPAMI.2010.128.

[30] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, *Logistic regression*. Springer, 2002.

[31] Y. Xiao, J. Wu, Z. Lin, and X. Zhao, "A deep learning-based multi-model ensemble method for cancer prediction," *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 1–9, 2018, doi: https://doi.org/10.1016/j.cmpb.2017.09.005.

[32] C. Goutte and E. Gaussier, "A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation," 2010.

[33] A. El-Sawy, H. EL-Bakry, and M. Loey, "CNN for Handwritten Arabic Digits Recognition Based on LeNet-5 BT - Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016," Cham, 2017, pp. 566–575.

[34] A. El-Sawy, M. Loey, and H. EL-Bakry, "Arabic Handwritten Characters Recognition Using Convolutional Neural Network," *WSEAS Transactions on Computer Research*, vol. 5, 2017, Accessed: Apr. 01, 2020. [Online]. Available: http://www.wseas.org/multimedia/journals/computerresearch/2017/a045818-075.php.

[35] N. E. M. Khalifa, M. H. N. Taha, A. E. Hassanien, and A. A. Hemedan, "Deep bacteria: robust deep learning data augmentation design for limited bacterial colony dataset," *International Journal of Reasoning-based Intelligent Systems*, 2019, doi: 10.1504/ijris.2019.102610.

[36] N. E. M. Khalifa, M. H. N. Taha, and A. E. Hassanien, "Aquarium Family Fish Species Identification System Using Deep Neural Networks," in *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2018*, 2018, pp. 347–356.

[37] N. E. M. Khalifa, M. H. N. Taha, D. Ezzat Ali, A. Slowik, and A. E. Hassanien, "Artificial Intelligence Technique for Gene Expression by Tumor RNA-Seq Data: A Novel Optimized Deep Learning Approach," *IEEE Access*, 2020, doi: 10.1109/access.2020.2970210.

[38] M. Loey, M. Naman, and H. Zayed, "Deep Transfer Learning in Diagnosing Leukemia in Blood Cells," *Computers*, vol. 9, no. 2, Art. no. 2, Jun. 2020, doi: 10.3390/computers9020029.

[39] J. G. A. Barbedo, "Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification," *Computers and Electronics in Agriculture*, vol. 153, pp. 46–53, Oct. 2018, doi: 10.1016/j.compag.2018.08.013.

[40] I. H. Sarker, A. S. M. Kayes, and P. Watters, "Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage," *Journal of Big Data*, vol. 6, no. 1, p. 57, Jul. 2019, doi: 10.1186/s40537-019-0219-y.

Highlights for Reviewers

Highlights:

- A hybrid deep and machine learning model proposed for face mask detection.

- The model can impede the Coronavirus transmission, specially COVID-19.

- Three face mask datasets have experimented with this research.

- The introduced model achieves high performance in the experimental study.