



ARTICLE

Fast and Accurate Detection of Masked Faces Using CNNs and LBPs

Sarah M. Alhammad¹, Doaa Sami Khafaga^{1,*}, Aya Y. Hamed², Osama El-Koumy³,
Ehab R. Mohamed³ and Khalid M. Hosny³

¹Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh, 11671, Saudi Arabia

²Computer Science Department, Sinai University, Arish, North Sinai, Egypt

³Department of Information Technology, Faculty of Computers and Informatics, Zagazig University, Zagazig, Egypt

*Corresponding Author: Doaa Sami Khafaga. Email: dskhafaga@pnu.edu.sa

Received: 07 April 2023 Accepted: 25 July 2023 Published: 09 November 2023

ABSTRACT

Face mask detection has several applications, including real-time surveillance, biometrics, etc. Identifying face masks is also helpful for crowd control and ensuring people wear them publicly. With monitoring personnel, it is impossible to ensure that people wear face masks; automated systems are a much superior option for face mask detection and monitoring. This paper introduces a simple and efficient approach for masked face detection. The architecture of the proposed approach is very straightforward; it combines deep learning and local binary patterns to extract features and classify them as masked or unmasked. The proposed system requires hardware with minimal power consumption compared to state-of-the-art deep learning algorithms. Our proposed system maintains two steps. At first, this work extracted the local features of an image by using a local binary pattern descriptor, and then we used deep learning to extract global features. The proposed approach has achieved excellent accuracy and high performance. The performance of the proposed method was tested on three benchmark datasets: the real-world masked faces dataset (RMFD), the simulated masked faces dataset (SMFD), and labeled faces in the wild (LFW). Performance metrics for the proposed technique were measured in terms of accuracy, precision, recall, and F1-score. Results indicated the efficiency of the proposed technique, providing accuracies of 99.86%, 99.98%, and 100% for RMFD, SMFD, and LFW, respectively. Moreover, the proposed method outperformed state-of-the-art deep learning methods in the recent bibliography for the same problem under study and on the same evaluation datasets.

KEYWORDS

Convolutional neural networks; face mask detection; local binary patterns; deep learning; computer vision; social protection; Keras; OpenCV; TensorFlow; Viola-Jones

1 Introduction

With the spread of COVID-19, millions of people have been infected worldwide, causing over 2 million deaths [1]. The infection can be transmitted through nasal secretions or saliva when an infected person sneezes or coughs since there is no particular treatment for this virus [2]. According



to the research work of Eikenberry et al. [3], wearing a mask can limit the danger of coronavirus transmission. Therefore, World Health Organization mandates wearing masks in public places; however, not everyone complies with the rules. Moreover, it is difficult to track down all the offenders in crowded places.

One of the most important fields is computer vision focuses on training machines to understand digital images or video sequences [4]. Therefore, computer vision could be used as an automatic real-time tool for monitoring people who do not comply with the rules. To this end, this work introduces an efficient computer vision approach to classifying human faces as masked and unmasked. Three standard benchmark datasets are used to assert the efficiency of the proposed method: the Real-World Masked Face Dataset (RMFD), the Simulated Masked Faces Dataset (SMFD), and the Labelled Faces in the Wild (LFW). Specifically, the proposed method exploits deep learning (DL) and the local descriptor local binary pattern (LBP) for face mask detection. First, the Haar cascade algorithm was employed to detect faces as one of the most popular and efficient face detection algorithms. Haar cascade algorithm was first introduced by Viola et al. [5] and has been commonly used to identify faces in images and real-time videos, proving its efficiency compared to alternative modern DL techniques [6].

The rest of this work is structured as follows. [Section 2](#) summarizes related work on masked face detection. [Section 3](#) presents materials and methods. Experimental results and discussion are included in [Section 4](#), while [Section 5](#) concludes the paper.

2 Related Work

Nowadays, due to the COVID-19 pandemic, masked face detection is a topic of interest. It should be noted that a search on Scopus on the term “masked face detection” returned 257 documents from 2019, when the pandemic started (February 2023). Due to the extensive number of related research, it is impossible to reference all of them; however, some were selected to be indicatively mentioned, focusing on recent publications that used the same datasets as the current work.

Verulkar et al. used viola jones for driver fatigue detection and could achieve an accuracy of 99.9% [7]. Sunar et al. could improve the Viola-Jones algorithm and introduce an approach to detect ovarian and breast cancer. He used the Viola-Jones to segment the images to find the region of interest. He could achieve an accuracy of 97% [8]. Another breast cancer detection technique built on top of Viola-Jones was proposed by [9]. Lu et al. [10] presented an algorithm for detecting spliced images for images that have humans, and then they used SVM to detect whether these images were fake or not. Many other applications use the Viola-Jones algorithm [11–13]. Anitha et al. [14] proposed a driver drowsiness technique using the Viola-Jones face detection algorithm to monitor the movements of drivers’ eyes for a pre-determined amount of time. If the drivers’ eyes were closing continuously, an alarm buzzed. Fatima et al. [15] used Viola-Jones for driver fatigue detection and reported an accuracy of 99.9%. Hussein et al. [16] improved the Viola-Jones algorithm and introduced an approach to detect ovarian and breast cancer. The authors used the Viola-Jones to segment the images and find the region of interest. Experimental results reported an accuracy of 97%. Another breast cancer detection technique built on top of Viola-Jones was also proposed by Magnuska et al. [17]. Hamad et al. [18] also proposed a face recognition system that uses the Viola-Jones algorithm to detect facial areas and could achieve high accuracy. The Viola-Jones algorithm has also been used in multiple alternative applications [19–23].

This work uses the Viola-Jones algorithm for face mask detection, combined with local binary patterns (LBPs) and a deep learning model. More specifically, this work introduces an efficient

algorithm that, for the first time to the authors' knowledge, combines LBPs and a convolutional neural network (CNN) to detect masked faces. Moreover, three well-known faced masked datasets, RMFD, SMFD, and LFW, are comparatively used to evaluate the performance of the proposed method to the problem under study.

Ryumina et al. [24] implemented a method for automatically recognizing the presence of a mask on a face. They combined Convolutional Neural Networks (CNN) with image histograms and tested their algorithm on the RMFD dataset. In [25], the authors developed a masked face detection system based on deep CNN (DCNN) models in conjunction with Extreme Learning Machines (ELM) as a classification technique. The RMDF dataset was also used in [26], where the authors employed the You Only Look Once (YOLOv3) object detection algorithm. The SMDF dataset was used in [27], where a CNN model and VGG16 were employed to detect masked and unmasked faces. All three datasets were employed in [28]; a hybrid model was proposed based on ResNet50 combined with classical machine learning (ML) algorithms. Rahman et al. [29] proposed a model developed by fine-tuning the pre-trained MobileNet. The LFW dataset was used in [30]; the authors introduced an implementation of FaceNet for masked face detection in extreme environments. Bae et al. [31] also trained their encoder with the images of LFW. Singh et al. [32] employed MobileNetV2 to determine the mask's presence. Sharma et al. [33] combined infrared temperature sensors with a CNN and used the RMFD dataset to detect masked faces. In [34], the authors used MobileNetV2 to determine mask presence. They used the SMFD dataset only during their experiment, with an accuracy of 99.016%. In [35], the authors presented a new masked face detection approach using hybrid machine learning techniques and achieved an accuracy of 97.32%.

Table 1 summarizes all information on the referenced research, including methods used, dataset, and performance results. It is obvious from the Table that the proposed method displays competitive performance compared to all previously referenced methods of the recent literature in all three datasets, as will be proved in the following sections.

Table 1: Comparative table of indicative related work on masked face detection of the recent literature using RMFD, SMFD, and LFW datasets

Authors, Ref., Year	Method	Dataset	Performance
Ryumina et al. [24], 2021	CNN (ResNet50) + image histograms	RMFD	98.12% precision, 99.89% recall
Agarwal et al. [25], 2022	DCNN (ResNet152) + ELM	RMFD	99.78% accuracy
Kaviya et al. [26], 2023	YOLOv3	RMFD	98.1% accuracy, 98.7% precision, 98% F1-score
Negi et al. [27], 2021	CNN	SMDF	100% precision, 98.59% recall, 99.29% F1-score
	VGG16		98.59% precision, 98.59% recall, 98.59% F1-score
Loey et al. [28], 2021	ResNet50 + ML models	RMFD	99.64% accuracy
		SMDF	99.49% accuracy
		LFW	100% accuracy
Rahman et al. [29], 2023	MobileNet	SMFD	99.61% accuracy, 99.6 precision, 99.63% recall
Wang et al. [30], 2023	FaceNet	LFW	99.76% accuracy
Bae et al. [31], 2023	LResNet50E-IR	LFW	99.17% accuracy
Singh et al. [32], 2022	MobileNetV2	SMDF	99.016% accuracy
Sharma et al. [33], 2022	CNN	RMFD	97.8% accuracy

(Continued)

Table 1 (continued)

Authors, Ref., Year	Method	Dataset	Performance
Proposed method	CNN + LBPs	RMFD	99.86% accuracy, 99.6% recall, 99.6% precision, 99.8% F1-score
		SMFD	99.98% accuracy, 99.8% recall, 99.8% precision, 99.9% F1-score
		LFW	100% accuracy, approx. 100% recall, 98.9% precision, 99.9% F1-score

3 Materials and Methods

3.1 Preliminaries

3.1.1 Viola-Jones Face Detector

The Viola-Jones detector is trained to recognize several object classes. It has also been applied to deal with the challenge of face detection and has proven its efficiency in real-time face detection. The algorithm uses a sliding window, which segments the image into sub-images. Then it searches these sub-images for Haar features, indicating the face characteristics, i.e., eyes, nose, mouth, etc. Fig. 1 shows the main Haar features.

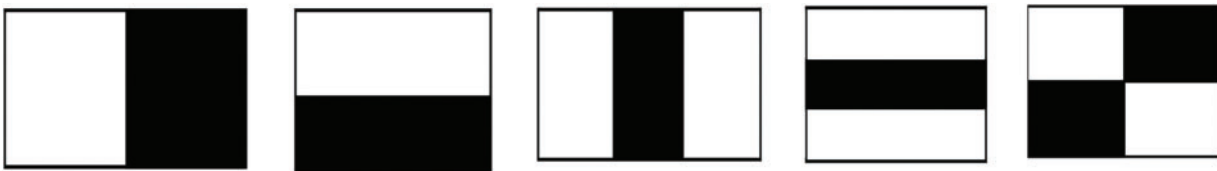


Figure 1: The main Haar features

3.1.2 Local Binary Pattern (LBP) Descriptor

LBP is a local visual descriptor that describes the local texture features of the input image. It was first introduced in 1994 [36,37]. The most important advantage of the LBP descriptor is that it is invariant to rotation and robust to illumination and transition. The traditional LBP works by using a window of size 3×3 . It compares the central pixel with each of the surrounding eight neighbors. If it is greater than or equal to the central pixel, it returns the value 1; otherwise, it returns the value 0. After that, the binary number is converted to its corresponding decimal value. Fig. 2 illustrates an example of the traditional LBP operator.

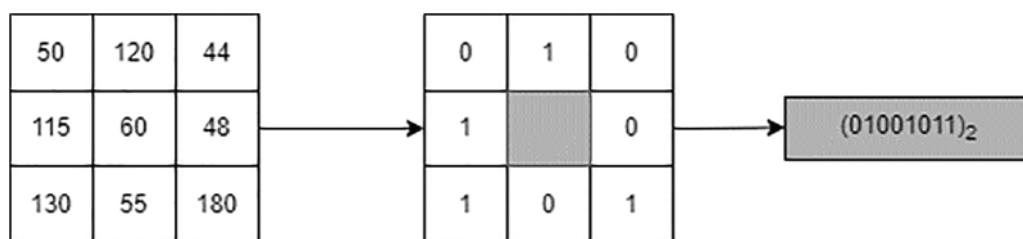


Figure 2: Uniform LBP operator

Therefore, LBPs work by thresholding the neighborhood of each pixel in the image and then converting them to binary numbers according to the following equations:

$$LBP(x, y) = \sum_{n=0}^7 C(I_n - I_{thrsh})2^n \tag{1}$$

where $C(r) = \begin{cases} 1 & r \geq 0 \\ 0 & \text{otherwise} \end{cases}$, and $I_{thrsh} = I(x, y)$ is the central pixel. Then the LBP histogram (LBPH) is used and calculated as follows:

$$LBPH(t) = \sum \delta\{t, LBP(x, y)\} \quad t = \{0, 1, \dots, 7\} \tag{2}$$

Ojala et al. [37] proposed an improved variation that could use neighbors of different sizes to take any radius and neighbors surrounding the central pixel, called a uniform local binary pattern. The uniform local pattern presents less feature vector length, and it is computed as follows:

$$LBP_{n,r}(I_{thrsh}) = \begin{cases} \sum_{n=0}^{n-1} C(I_n - I_{thrsh}) 2^n & \text{if } U(LBP_{n,r}) > 2 \\ n + 1 & \text{otherwise} \end{cases} \tag{3}$$

$$LBPH_n = \sum f\{I(x, y) = n\} \quad n = \{0, 1, \dots, n - 1\}$$

where r is the radius, and n is the number of neighbors. Fig. 3 shows the resulting coded image data.

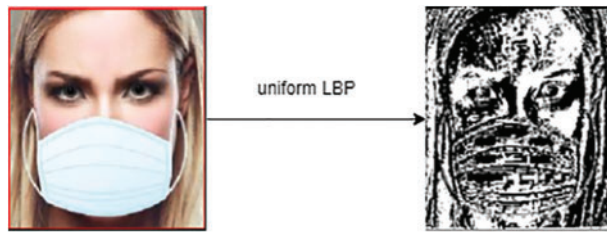


Figure 3: Uniform LBP operator

3.1.3 Convolutional Neural Network (CNN) Model

The convolutional neural network is an artificial neural network widely used in image recognition. CNN models have proven their efficiency in computer vision, which motivated us to deploy them in this work. The motivation for using the CNN model for feature description is the deep neural network’s ability to extract high-level features that are the main points to distinguish images. Our proposed model includes the following: input layer, convolution layer, max pooling, average pooling layer, fully connected layer, and output layer.

3.2 The Proposed Methodology

The proposed method comprises two main steps: (1) the first step is face detection using the Viola-Jones detector; (2) the second step is the masked face classification using a new CNN model and LBP descriptor. The proposed method uses established techniques that have proven their efficiency and combined them for the first time towards masked face detection.

The Viola-Jones descriptor puts a bounding box around the face region [37]. Then, a cropping technique is applied to each box to identify the region of interest, as shown in Fig. 4. After that, the LBP algorithm is used to extract local features of the face area to get the LBP image which is used

as an input image to the CNN model. The following section discusses the proposed CNN model in detail.

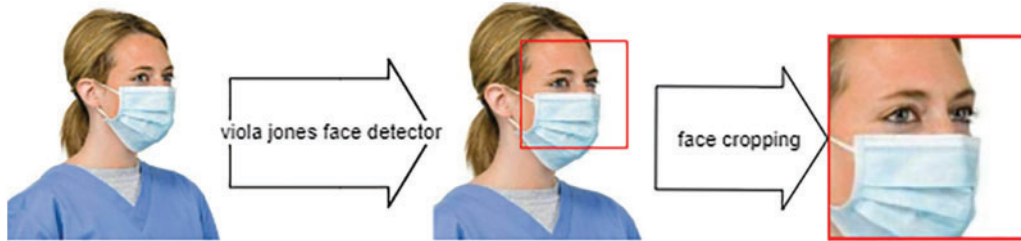


Figure 4: Generation of the region of interest

3.2.1 Features Extraction and Classification

The architecture of the CNN model proposed in this work is shown in Fig. 5. It comprises six convolution layers, six-max pooling layers, one average pooling layer, one flattened layer, finally dense layer, two fully connected layers, and one output layer. As illustrated in Table 2, the model is simple; the total number of parameters is 1,573,794, out of which 446,306 are non-trainable; hence, the number of trainable parameters utilized is 1,127,488, which is suitable for environments with limited resources such as Raspberry Pi. The number of parameters is small when compared to the state of art approaches, as shown in Table 3.

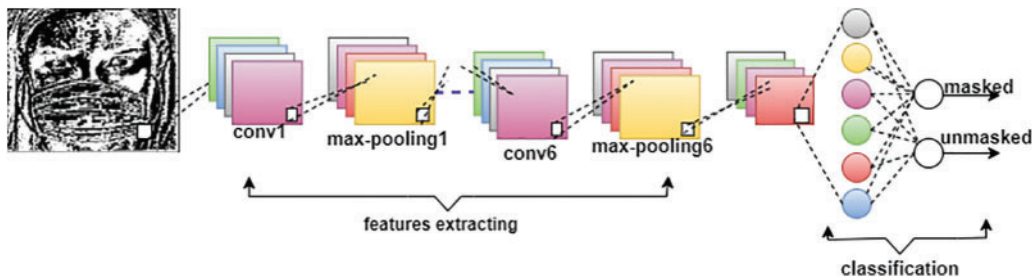


Figure 5: The proposed CNN model structure

Table 2: The proposed model activation shape, activation size, and hyperparameters

Layers	Activation shape	Activation size	Number of parameters
Input layer	(224,224,1)	50,176	0
Conv1	(222,222,16)	788,544	448
Max pool1	(111,111,16)	197,136	0
Conv2	(109,109,32)	380,192	4640
Max pool2	(54,54,32)	93,312	0
Conv3	(52,52,64)	173,056	18,496
Max pool3	(26,26,64)	43,264	0
Conv4	(24,24,128)	73,728	73,856
Max pool4	(12,12,128)	18,432	0

(Continued)

Table 2 (continued)

Layers	Activation shape	Activation size	Number of parameters
Conv5	(10,10,256)	25,600	295,168
Max pool5	(5,5,256)	6,400	0
Conv6	(3,3,512)	4,608	1,180,160
Max pool6	(1,1,512)	512	0
Average pooling 2D	512	512	0
Fully connected	512	512	0
Output layer	2	2	1026
Total number of parameters			1,573,794

Table 3: Comparison of hyperparameters to other approaches

Authors, Ref., Year	Method	#parameters
Ryumina et al. [24], 2021	CNN (ResNet50) + image histograms	–
Agarwal et al. [25], 2022	DCNN (ResNet152) + ELM	4,653,2048
Kaviya et al. [26], 2023	YOLOv3	–
Negi et al. [27], 2021	CNN	2,181,778
	VGG16	14,780,610
Loey et al. [28], 2021	ResNet50 + ML models	–
Rahman et al. [29], 2023	MobileNet	3,360,709
Wang et al. [30], 2023	FaceNet	–
Bae et al. [31], 2023	LResNet50E-IR	–
Singh et al. [32], 2022	MobileNetV2	2,257,984
Sharma et al. [33], 2022	CNN	–
Proposed method	CNN + LBPs	1,573,794

The process is as follows:

- The size of the input image is 224×224 . At first, the local features are extracted from the input image using the LBP descriptor to generate the LBP-coded image of the same size. Then the LBP-coded image is passed as an input layer to the CNN model.
- The convolution neural network model extracts the distinctive features using a convolution layer with kernel size 3×3 . This work used six convolution layers, each followed by one max pooling layer to shrink the number of features; the filter size of the max pooling layers is 2×2 . Then, one average pooling layer was added to minimize the total number of extracted features. Next, the flattened layer was used to get a 1D feature vector, and finally, the feature vector was passed to the fully connected layer for classification.

The steps of the proposed methodology can be summarized as follows:

- Application of Viola-Jones to detect face region.
- Cropping of the region of interest.

- Application of local binary pattern descriptor to get LBP-coded image.
- Applying a CNN model to an LBP-coded image to classify the faces as masked and unmasked.

3.2.2 System Architecture

In this work, the Raspberry Pi 4 computer model B with 1 GB RAM, which runs the Debian operating system, was used for all experiments; the tools used are summarized in Table 4. Raspberry Pi 4 was chosen because of its small size and high performance. When using these nano-devices, size is of great importance. The proposed model is suitable for raspberry pi's abilities. The dataset is loaded and then fed to the CNN model for training.

Table 4: Details of the tools used

Name	Configuration
Operating system	Debian
Programming language	Python
Libraries	OpenCV 4.1.0, Keras, and TensorFlow
Optimizer	Adam
Batch size	32
Epoch	16

4 Experimental Results and Discussion

The proposed technique was evaluated on three benchmark datasets: the Real-World Masked Face Dataset (RMFD), the Simulated Masked Faces Dataset (SMFD), and the Labelled Faces in the Wild (LFW). Experimental results and their analysis on each dataset occur in what follows.

4.1 Datasets

These datasets have been chosen because images in each dataset vary in background, head scale, translation, facial emotions, and image lighting.

Researchers from Wuhan University gathered and developed the RMFD dataset [38]. It contained 5,000 masked faces and 90,000 unmasked faces for 252 different identities. Fig. 6 shows some indicative samples of the RMFD dataset. Since this dataset was collected from the internet, it contains errors, blurry images, and duplicated images. Our experiment used 755 images for masked faces and 754 for unmasked faces. The same subgroup was introduced in [33].



Figure 6: Indicative images of the RMFD dataset [38]

The second dataset was SMFD [39,40]. It contained 1,570 images for different people, 758 images for people with simulated masked faces, and 758 images for unmasked faces. Samples of this dataset are shown in Fig. 7. Both RMFD and SMFD were used for training, validation, and testing.



Figure 7: Indicative images of the SMFD dataset [40]

The third dataset was LFW [41–43]. LFW dataset contained 13,000 simulated masked faces for different celebrities around the world. The LFW dataset was used only for the testing phase, and the proposed model was never trained on it. Fig. 8 shows some indicative samples of the LFW dataset.



Figure 8: Indicative images of the LFW dataset [42]

4.2 Preliminary CNN Model Results

In the first stage of the experiments, the proposed CNN model is tested on the data without using the LBP descriptor. In the second stage, the LBP-coded images were added to highlight their contribution and show that their combined use can further improve the results.

Therefore, at this first step, the extracted ROI (face area) from the input images using Viola-Jones [21] was applied to the CNN model for the RMFD and SMFD datasets. The training set was used to train the model. A ratio of 75:25 for training and testing (50% training, 25% validation, and 25% testing) was used. Fig. 9 shows the accuracy and loss rate results during the training and validation. As can be seen from the figure, the training accuracy rate reached 100% at epoch 16. In order to avoid overfitting, we used early stopping technique in which the training process is stopped when the performance on validation set stops improving to prevent the model from continuing to learn the noise in training data.

Then, the testing set was used to evaluate the model's performance. Accuracy and other performance metrics were calculated to evaluate the proposed innovative approach.

More specifically, the evaluation metrics that were considered were the following:

$$Accuracy = \frac{TP + FP}{TP + FP + TN + FN} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (7)$$

The TP, TN, FP, and FN refer to the true positive, true negative, false positive, and false negative, respectively. Table 5 shows the evaluation of the proposed CNN model over the three datasets.

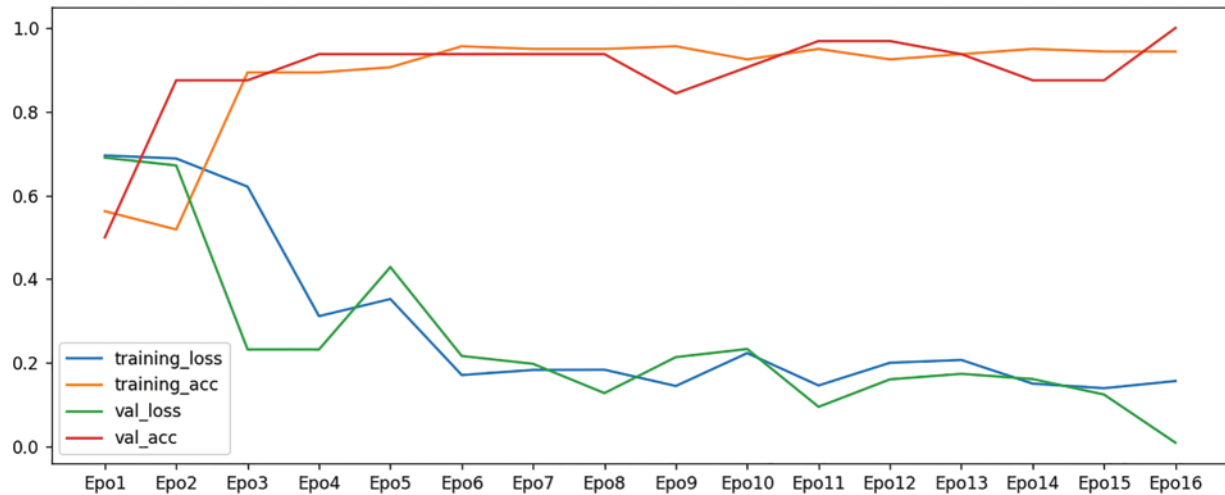


Figure 9: Results of detection accuracy and loss rates during training and validation

Table 5: Evaluation results of the proposed CNN model over the three datasets

Dataset	Recall	Precision	F1-score	Accuracy (%)
RMFD	0.989	0.990	0.990	99.016
SMFD	0.994	0.994	0.995	99.48
LFW	0.996	0.991	0.997	99.65

The performance was evaluated for the three datasets, and the confusion matrix and accuracy were calculated. The highest reported accuracies were 99.016% for the RMFD dataset, 99.48% for the SMFD dataset, and 99.65% for the LFW dataset.

Despite the high accuracy, further enhancement was pursued. Obtained results motivated the authors to think of LBPs as excellent descriptors able to extract local features from the image and detect the edges. LBPs have proved their efficiency in extracting distinctive features [44,45]. Therefore, the LBP descriptor was applied to the input images, and a second experimental trial took place to get LBP-coded images towards enhanced classification accuracy, as follows.

4.3 Proposed LBP + CNN Model Results

In the second step of the experiments, the LBP-coded images were used as an input layer to the CNN model. Results are included in Table 6.

Table 6: Evaluation of the proposed methodology (LBPs + CNN) over the three datasets

Dataset	Recall	Precision	F1-score	Accuracy (%)
RMFD	0.996	0.996	0.998	99.86
SMFD	0.998	0.998	0.999	99.98
LFW	0.991	0.989	0.999	100.00

The combination of LBPs with the CNN resulted in improved performance. The reported accuracy for the RMFD dataset was 99.86%, 99.98% for the SMFD dataset, and 100% for the LFW dataset.

A batch size of 35 and an Adam optimizer were utilized in the classifiers. The model was trained for 30 epochs. The initial learning rate was 0.04, the learning accuracy reached 100% at epoch 16, and the training was finished at the 30th epoch.

The obtained results justify the great utility of LBPs to uniquely describe an image's texture in contrast to the common RGB images. In this context, the CNN that receives the LBP images as inputs can generate even more useful features using a smaller architecture depth. Obtained experimental results are also highly competitive compared to other state-of-the-art CNN models of the recent literature evaluated on the same dataset.

5 Conclusions

Since COVID-19 is one of the most difficulties humanity has ever encountered, wearing a face mask can effectively stop the spread of this deadly infection. This work proposes a new computer vision method for detecting masked faces, combining LBP with a CNN model. The motivation was to develop a simple technique to run on nano-devices like Raspberry Pi. The proposed CNN model was enhanced by using local feature descriptors such as LBP and, thus, achieved high performance because CNN algorithms use global features of the images. More specifically, the Viola-Jones algorithm was adopted to extract the face area from the input images. The method was experimentally evaluated on three different benchmark datasets. The evaluation results demonstrated that the proposed method outperformed recently proposed methods of the literature for the same problem under study and on the same datasets. The reported accuracy results of the proposed method were 99.86%, 99.98%, and 100% for the RMFD, SMFD, and LFW datasets, respectively.

Future work will focus on training the proposed CNN model to detect incorrectly worn masks. The proposed method can also be extended to regular occlusion objects like sunglasses and scarves. Environmental noise, such as lighting conditions, could also be investigated for detecting and recognizing masked faces under low lighting conditions. Moreover, current trends in computer vision, such as deep domain adaptation methods, could be investigated to see how they could boost the model's performance.

Acknowledgement: This project is funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project Number (PNURSP2023R442), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Funding Statement: Princess Nourah bint Abdulrahman University Researchers Supporting Project Number (PNURSP2023R442), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author Contributions: Conceptualization, D.S.K, S.M.A, K.M.H, O.E, E.R.M, and A.Y.H; methodology, D.S.K, S.M.A, K.M.H, and A.Y.H; software, D.S.K, S.M.A, K.M.H, and A.Y.H; validation, D.S.K, S.M.A, K.M.H, O.E, E.R.M, and A.Y.H; formal analysis, D.S.K, S.M.A, K.M.H, O.E, E.R.M, and A.Y.H; investigation, D.S.K, S.M.A, K.M.H, O.E, E.R.M, and A.Y.H; resources, D.S.K, S.M.A, K.M.H, O.E, E.R.M, and A.Y.H; data curation, D.S.K, S.M.A, K.M.H; writing—original draft preparation, D.S.K, S.M.A, K.M.H, O.E, E.R.M, and A.Y.H; writing—review and editing, D.S.K, S.M.A, E.V and G.A.P; visualization, K.M.H, O.E, E.R.M, and A.Y.H; supervision, E.R.M; project administration, E.R.M; funding acquisition, G.A.P. All authors have read and agreed to the published version of the manuscript.

Availability of Data and Materials: The Real-World Masked Face Dataset (RMFD) is available online at <https://github.com/X-zhangyang/Real-World-Masked-Face-Dataset> (accessed on Nov 11, 2022) [36]. The Simulated Masked Faces Dataset (SMFD) is available online at <https://github.com/prajnasb/observations> (accessed on Nov 11, 2022) [38]. The Labelled Faces in the Wild (LFW) is available online at <https://www.kaggle.com/muhammeddalkran/lfw-simulated-masked-face-dataset> (accessed on Nov 11, 2022) [40].

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] Johns Hopkins University (JHU) & Medicine, “COVID-19 dashboard by the center for systems science and engineering (CSSE),” 2023. [Online]. Available: <https://coronavirus.jhu.edu/map.html>
- [2] Q. Cai, M. Yang, D. Liu, J. Chen, D. Shu *et al.*, “Experimental treatment with favipiravir for COVID-19: An open-label control study,” *Engineering*, vol. 6, no. 10, pp. 1192–1198, 2020.
- [3] S. E. Eikenberry, M. Mancuso, E. Iboi, T. Phan, K. Eikenberry *et al.*, “To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic,” *Infectious Disease Modelling*, vol. 5, pp. 293–308, 2020.
- [4] N. Doulamis, A. Doulamis and E. Protopapadakis, “Deep learning for computer vision: A brief review,” *Computational Intelligence and Neuroscience*, vol. 2018, no. 7068349, pp. 1–13, 2018.
- [5] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proc. of the 2001 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, CVPR 2001*, Kauai HI, USA, pp. I, 2001.
- [6] J. M. Al-Tuwaijari and S. A. Shaker, “Face detection system based Viola-Jones algorithm,” in *2020 6th Int. Engineering Conf. “Sustainable Technology and Development” (IEC)*, Erbil, Iraq, pp. 211–215, 2020.
- [7] A. Verulkar and K. Bhurchandi, “Auto-luminance-based face image recognition system,” in *Proc. of First Int. Conf. on Computational Electronics for Wireless Communications Springer*, Singapore, vol. 329, pp. 553–565, 2022.
- [8] S. Sunar, S. K. Tripathi, U. Tiwari and H. Srivastava, “A comparative study on face recognition AI robot,” in *Proc. Second Doctoral Symp. on Computational Intelligence*, vol. 1374, pp. 211–221, 2022.
- [9] G. V. Sai Prasanna, K. Pavani and M. Kumar Singh, “Spliced images detection by using Viola-Jones algorithms method,” *Materials Today: Proceedings*, vol. 51, pp. 924–927, 2022.
- [10] C. T. Lu, C. W. Su, H. L. Jiang and Y. Y. Lu, “An interactive greeting system using convolutional neural networks for emotion recognition,” *Entertainment Computing*, vol. 40, pp. 100452, 2022.
- [11] S. Sharma, L. Raja, V. Bhatnagar, D. Sharma and S. N. Bhagirath, “Hybrid HOG-SVM encrypted face detection and recognition model,” *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 25, pp. 205–218, 2022.
- [12] K. M. Hosny, M. Abd Elaziz and M. M. Darwish, “Color face recognition using novel fractional-order multi-channel exponent moments,” *Neural Computing and Applications*, vol. 33, pp. 5419–5435, 2021.

- [13] K. M. Hosny and M. A. Elaziz, "Face recognition using exact Gaussian-Hermit moments," in *Recent Advances in Computer Vision. Studies in Computational Intelligence*, vol. 804. Cham: Springer, pp. 169–187, 2019.
- [14] J. Anitha, G. Mani and K. Venkata Rao, "Driver drowsiness detection using Viola Jones algorithm," in *Smart Intelligent Computing and Applications. Smart Innovation, Systems, and Technologies*, vol. 159. Singapore: Springer, pp. 583–592, 2020.
- [15] B. Fatima, A. R. Shahid, S. Ziauddin, A. A. Safi and H. Ramzan, "Driver fatigue detection using Viola Jones and principal component analysis," *Applied Artificial Intelligence*, vol. 34, pp. 456–483, 2020.
- [16] I. J. Hussein, M. A. Burhanuddin, M. A. Mohammed, M. Elhoseny, B. Garcia-Zapirain *et al.*, "Fully automatic segmentation of gynecological abnormality using a new Viola-Jones model," *Computers, Materials & Continua*, vol. 66, no. 3, pp. 3161–3182, 2021.
- [17] Z. A. Magnuska, B. Theek, M. Darguzyte, M. Palmowski, E. Stickeler *et al.*, "Influence of the computer-aided decision support system design on ultrasound-based breast cancer classification," *Cancers*, vol. 14, pp. 1–20, 2022.
- [18] A. Y. Hamad, K. M. Hosny, O. Elkomy and E. R. Mohamed, "Fast and accurate face recognition system using MORSCMS-LBP on embedded circuits," *PeerJ Computer Science*, vol. 8, pp. 1–19, 2022.
- [19] K. D. Ismael and S. Irina, "Face recognition using Viola-Jones depending on python," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 3, pp. 1513–1521, 2020.
- [20] M. F. Hirzi, S. Efendi and R. W. Sembiring, "Literature study of face recognition using the Viola-Jones algorithm," in *Int. Conf. on Artificial Intelligence and Mechatronics Systems (AIMS)*, Bandung, Indonesia, pp. 1–6, 2021.
- [21] M. Ghosh, T. Sarkar, D. Chokhani and A. Dey, "Face detection and extraction using Viola-Jones algorithm," in *Computational Advancement in Communication, Circuits, and Systems*, Singapore, Springer, vol. 786, pp. 93–107, 2022.
- [22] Y. Tang, Z. Pan, W. Pedrycz, F. Ren and X. Song, "Viewpoint-based kernel fuzzy clustering with weight information granules," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 7, no. 2, pp. 342–356, 2023.
- [23] Y. M. Tang, L. Zhang, G. Q. Bao, F. J. Ren and W. Pedrycz, "Symmetric implicational algorithm derived from intuitionistic fuzzy entropy," *Iranian Journal of Fuzzy Systems*, vol. 19, no. 4, pp. 27–44, 2020.
- [24] E. Ryumina, D. Ryumin, D. Ivanko and A. Karpov, "A novel method for protective face mask detection using convolutional neural networks and image histograms," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIV-2/W1, pp. 177–182, 2021.
- [25] C. Agarwal, P. Itondia and A. Mishra, "A novel DCNN-ELM hybrid framework for face mask detection," *Intelligent Systems with Applications*, vol. 17, pp. 200175, 2023.
- [26] P. Kaviya, P. Chitra and B. Selvakumar, "A unified framework for monitoring social distancing and face mask wearing using deep learning," *Procedia Computer Science*, vol. 218, pp. 1561–1570, 2023.
- [27] A. Negi, K. Kumar, P. Chauhan and R. S. Rajput, "Deep neural architecture for face mask detection on simulated masked face dataset against COVID-19 pandemic," in *Int. Conf. on Computing, Communication and Intelligent Systems (ICCCIS)*, Greater Noida, India, pp. 595–600, 2021.
- [28] M. Loey, G. Manogaran, M. H. N. Taha and N. E. M. Khalifa, "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic," *Measurement*, vol. 167, no. 5, pp. 108288, 2021.
- [29] Md H. Rahman, Mir K. Ara Jannat, Md S. Islam, G. Grossi, S. Bursic *et al.*, "Real-time face mask position recognition system based on mobilenet model," *Smart Health*, vol. 28, pp. 100382, 2023.
- [30] Y. Wang, Y. Li and H. Zou, "Masked face recognition system based on attention mechanism," *Information- an International Interdisciplinary Journal*, vol. 14, no. 2, pp. 87, 2023.
- [31] G. Bae, M. de La Gorce, T. Baltrusaitis, Ch. Hewitt, D. Chen *et al.*, "DigiFace-1M: 1 million digital face images for face recognition," in *Proc. of the CVF Winter Conf. on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, pp. 3526–3535, 2023.

- [32] J. Singh, L. Gupta, R. Kushwaha, T. Varshney and A. Chauhan, “Real-time face mask detection and analysis system,” in *Proc. of Data Analytics and Management*, Singapore: Springer, vol. 90, pp. 11–19, 2022.
- [33] A. Sharma, J. Dhingra and P. Dawar, “A check on who protocol implementation for COVID-19 using IoT,” in *Internet of Things and its Applications. EAI/Springer Innovations in Communication and Computing*, Cham: Springer, pp. 63–79, 2022.
- [34] S. K. Dey, A. Howlader and C. Deb, “MobileNet mask: A multi-phase face mask detection model to prevent person-to-person transmission of SARS-CoV-2,” in *Proc. of Int. Conf. on Trends in Computational and Cognitive Engineering*, Singapore, Springer, vol. 1309, pp. 603–613, 2021.
- [35] B. Wang, Y. Zhao and C. L. Philip Chen, “Hybrid transfer learning and broad learning system for wearing mask detection in the COVID-19 era,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, no. 5009612, pp. 1–12, 2021.
- [36] T. Ojala, M. Pietikainen and D. Harwood, “Performance evaluation of texture measures with classification based on kullback discrimination of distributions,” in *Proc. of 12th Int. Conf. on Pattern Recognition*, Jerusalem, Israel, vol. 1, pp. 582–585, 1994.
- [37] T. Ojala, M. Pietikainen and D. Harwood, “A comparative study of texture measures with classification based on featured distributions,” *Pattern Recognition*, vol. 29, pp. 51–59, 1996.
- [38] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong *et al.*, “Real-world masked face dataset (RMFD),” 2020. [Online]. Available: <https://github.com/X-zhangyang/Real-World-Masked-Face-Dataset>
- [39] G. J. Chowdary, N. S. Punn, S. K. Sonbhadra and S. Agarwal, “Face mask detection using transfer learning of inceptionV3,” in *Lecture Notes in Computer Science*, Cham, Springer, vol. 2, pp. 81–90, 2020.
- [40] Prajnasb Prajna Bhandary, “Simulated masked faces dataset (SMFD),” 2020. [Online]. Available: <https://github.com/prajnasb/observations>
- [41] E. Learned-Miller, G. Huang, A. Roy Chowdhury, H. Li and G. Hua, “Labeled faces in the wild: A survey,” in *Advances in Face Detection and Facial Image Analysis*, Cham: Springer International Publishing, pp. 189–248, 2016.
- [42] M. Dalkiran, “LFW simulated masked face dataset,” 2020. [Online]. Available: <https://www.kaggle.com/datasets/muhammeddalkran/lfw-simulated-masked-face-dataset>
- [43] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong *et al.*, “Masked face recognition dataset and application,” *IEEE Transactions on Biometrics, Behavior and Identity Science*, vol. 5, no. 2, pp. 298–304, 2020.
- [44] S. Banerji, A. Verma and C. Liu, “LBP and color descriptors for image classification, cross disciplinary biometric systems,” *Intelligent Systems Reference Library*, vol. 37, pp. 205–225, 2012.
- [45] H. P. Truong, T. P. Nguyen and Y. G. Kim, “Statistical binary patterns for facial feature representation,” *Applied Intelligence*, vol. 52, pp. 1893–1912, 2022.