applied sciences

MDPI

Article

Automatic Face Recognition System Using Deep Convolutional Mixer Architecture and AdaBoost Classifier

Qaisar Abbas[1,*], Talal Al-Balawi[1], Ganeshkumar Perumal[1], M Emre Celebi[2]

[1] College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh; qaabbas@imamu.edu.sa
[2] Department of Computer Science and Engineering, University of Central Arkansas, 201 Donaghey Ave., Conway, AR, 72035, USA; ecelebi@uca.edu.
* Correspondence: qaabbas@imamu.edu.sa; Tel.: 00966537014011

Abstract: In recent years, advances in deep learning (DL) techniques for video analysis have developed to solve the problem of real-time processing. Automated face recognition in the runtime environment has become necessary in video surveillance systems for urban security. This is a difficult task due to face occlusion, which makes it hard to capture effective features. Existing work focuses on improving performance while ignoring issues like a small dataset, high computational complexity, and a lack of lightweight and efficient feature descriptors. In this paper, face recognition (FR) using a Convolutional mixer (AFR-Conv) algorithm is developed to handle face occlusion problems. A novel AFR-Conv architecture is designed by assigning priority-based weight to the different face patches along with residual connections and an AdaBoost classifier for automatically recognizing human faces. The AFR-Conv also leverages the strengths of pre-trained CNNs by extracting features using ResNet-50, Inception-v3, and DenseNet-161. The AdaBoost classifier combines these features' weighted votes to predict labels for testing images. To develop this system, we use the data augmentation method to enhance the number of datasets using human face images. The AFR-Conv method is then used to extract robust features from images. Finally, to recognize human identity, an AdaBoost classifier is utilized. For the training and evaluation of the AFR-Conv model, a set of face images is collected from online data sources. The experimental results of the AFR-Conv approach are presented in terms of precision (PR), recall (RE), detection accuracy (DA), and F1-score metrics. Particularly, the proposed approach attains 95.5% PR, 97.6% RE, 97.5% DA, and 98.5% of F1-score on 8,500 face images. The experimental results show that our proposed scheme outperforms advanced methods for face classification. The AFR-Conv model code is freely available on GitHub (https://github.com/Qaisar256/AFR-ConvMixer) for the scientific community.

Keywords: Intelligent Systems, Internet of Things; Covid-19; Face mask; Computer vision; Video analysis; Face occlusion; Deep learning; Transfer learning; Convolutional neural network; ConvMixer model

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Lastname

Received: date
Revised: date
Accepted: date
Published: date

1. Introduction

The current epidemic of COVID-19 spreads all over the world, forcing people to cover most of their faces with masks to contain the pandemic. Despite significant progress in facial detection and recognition research during the previous decade [1], present facial recognition systems (FRS) need to be more precise and robust to be fully implemented in high-security contexts. FR has been a significant topic in science for the past three decades, with numerous vital activities related to identity verification, recognizing criminal activities, surveillance, and scientific study. Face recognition integrated with Internet of Things (IoT) devices is a cutting-edge application of artificial intelligence. These intelligent systems enhance security, convenience, and personalization by cap-

turing and analyzing facial features. The potential applications are diverse, from smart doorbells granting access to authorized individuals to personalized experiences delivered by intelligent mirrors. However, as with any technology, responsible implementation is critical. Striking the right balance between innovation and privacy protection is vital to ensuring these systems benefit society without compromising individual rights. However, traditional face detection and recognition algorithms need to improve their intelligence. Hence, there is an urgent need for modern face detection and recognition methods that effectively deal with occluded faces and result in higher detection and recognition accuracies. Masked Face Recognition (MFR) is a unique occlusion-based FR application. In contrast to regular occlusion FR, MFR has three significant challenges [2]. To begin with, there is a need for larger face datasets with masks. Second, masks completely occlude the mouth and nasal characteristics, reducing the effectiveness of facial feature extraction. So, it is challenging to identify a human when an act covers the face. At the same time, two unique instances are difficult to tackle using existing deep learning (DL) methods, such as face masks used for training and non-face covers for testing, and vice versa. Yet, under some unique circumstances, the two scenarios are critical. During the COVID-19 outbreak, for example, standard FR systems could not distinguish faces wearing masks. The authorities have a lot of occlusion images of the suspects, but they don't have any clear face images.

The concept of deep learning (DL) [3–7] is widely utilized in many applications. However, because of the long time required for network training, using DL in a real-time environment was challenging. Since then, the recent advances in DL approaches proposed in [4], [5], [6], and [7] have motivated other authors to use the advancement of software and hardware in a parallel computing environment. Convolutional Neural Networks (CNN), Deep Belief Networks (DBN), Restricted Boltzmann Machines (RBM), Recursive Neural Networks (RNN), and Stack-Based Auto-Encoders (SAE) are examples of deep learning architectures that are effectively used in a variety of applications such as natural language processing (NLP), bioinformatics, and computer vision [8], [9]. Deep learning-based techniques have benefited face detection, recognition, and forecasting applications [10], [11].

Several existing FAR systems used the classic computer vision techniques highlighted in Section 2. Regarding their feature extraction approaches, these strategies have limitations. All of these works use feature extractors that were handcrafted. As a result, the extracted image features reflect the variations between the natural and occluded face images in the spatial and frequency domains. In a few research studies, we noticed that deep learning (DL) methods are used to construct FRS systems. The FRS system used AlexNet and VGG networks in previous DL-based ways to extract features and recognize human images. The AlexNet model, on the other hand, has only eight layers, making it a shallow model. As a result of facial occlusion, extracting robust characteristics from human face images could be more effective. Furthermore, achieving decent accuracy is time-consuming, which could be more efficient for instantaneous FRS applications. Compared to the newer DL model, the VGG network model has a vanishing gradient problem and is also extremely slow. The current DL-based convolutional neural network (CNN) model includes many parameters and calculations requiring complex hardware. A learning-based technique called a ConvMixer [33] was recently introduced that uses image patches along with a convolution-based architecture instead of a transformer-based architecture. The authors use depthwise convolution followed by pointwise convolution as their main ConvMixer layer, which is repeated depth times. In practice, the ConvMixer architecture gives better validation accuracy than a basic CNN model with four times fewer parameters. The ConvMixer has been successfully used to extract features from image-based recognition systems. These ConvMixer models outperform several earlier transfer learning (TL) models and classic handcrafted feature extraction

methods. Accordingly, this paper proposes a new FAR method for FR systems based on a residual connection in the ConvMixer architecture with AdaBoost.

This study provides a novel technique for dealing with challenges caused by facial occlusions and variations in facial expression. These limitations exist in recognizing human faces [10], [11]. Figure 1 shows a visual example of different face occlusions, including COVID-19 masks that make it difficult to recognize human faces automatically. This paper proposes a deep learning methodology (ENSEMBLE-FRO) for video analysis and surveillance systems, especially in partially occluded environments where the look is invisible. The proposed ENSEMBLE-FRO comprises three pre-trained DL architectures: ResNeXt-50, Inception-v3, and DenseNet-161. Using an augmentation method, the authors create a synthetic face mask evaluation dataset using many prominent public verification datasets, including LFW, CALFW, CPLFW, and CFP. The Real-World Masked Face Dataset (RMFD) is used in addition to the synthesized versions of typical FR testing datasets. Several performance metrics are used to assess the performance of the proposed technique, such as precision, recall, accuracy, and F1-score.

1.1. Research Motivations

The problem addressed in this work is developing an efficient and accurate face recognition system to overcome challenges posed by face occlusion and improve recognition performance in real-world scenarios. Face occlusion, such as partial face coverage due to accessories, obstructions, or poor lighting conditions, is typical in video surveillance and urban security applications. Existing face recognition algorithms often struggle to handle these challenging conditions, leading to reduced accuracy and reliability due to wearing face masks. The motivations for developing a facial recognition system for people with and without face masks during the COVID-19 pandemic are:

1) The COVID-19 pandemic has underscored the importance of minimizing physical contact and maintaining hygiene. Implementing a contactless identification system like facial recognition can help reduce the risk of virus transmission through shared touchpoints, such as fingerprint scanners.
2) The widespread adoption of face masks as a preventive measure presents a challenge for traditional facial recognition systems designed for unmasked faces. Developing a system that can accurately recognize individuals both with and without face masks addresses this compliance monitoring need.
3) The pandemic has accelerated the adoption of face masks as a new societal norm. A facial recognition system capable of functioning effectively in the presence of face masks aligns with these evolving norms and ensures seamless integration into daily activities.
4) Surveillance and security applications benefit from accurate facial recognition, especially in crowded places like airports, public transportation, and essential service facilities. A system that can recognize faces despite masks contributes to enhanced public safety.
5) Traditional face recognition systems face accuracy and reliability issues when dealing with partial face coverage due to masks. This motivates the development of innovative solutions that can mitigate the negative impact of covers on recognition performance.
6) The pandemic has generated much data regarding masked and unmasked faces. Leveraging this data for research and development purposes offers a unique opportunity to create more robust and effective facial recognition systems.
7) Addressing the challenges posed by face masks in facial recognition requires innovation. Developing a system that can accurately recognize faces under diverse conditions reflects advancements in computer vision and deep learning techniques.

The motivation stems from the need to adapt facial recognition technology to the current global context, ensuring safety, accuracy, and seamless integration with public health measures during and beyond the COVID-19 pandemic. The objective of this research is to propose a solution that leverages the ConvMixer architecture specifically designed for face recognition, along with the integration of an AdaBoost classifier, to handle face occlusion effectively, enhance feature representation, and achieve superior recognition accuracy compared to other state-of-the-art deep learning algorithms. The study aims to evaluate the proposed system's performance using benchmark datasets and validate its generalizability and efficiency for real-time deployment in face recognition applications. The goal is to provide a robust and practical face recognition solution that recognizes human faces even under challenging real-world conditions, contributing to advancing urban security and video surveillance technologies.

1.2. Major Contributions

The major contribution of this work lies in the development of a novel ConvMixer and AdaBoost-based face recognition system that effectively addresses face occlusion challenges and outperforms existing deep learning algorithms. Its potential for transfer learning and real-world applicability make it a valuable solution for enhancing face recognition accuracy and reliability in critical surveillance applications. The proposed FAR-Conv approach differs from previous methodologies in the following four aspects.

1. A new face recognition system (AFR) method to handle issues of face occlusion based on a residual connection and ConvMixer with AdaBoost is developed in this study to addresses data limitation, computational cost, and the lack of a lightweight and efficient feature descriptor.
2. We address the challenges of AFR in two different scenarios: utilizing masked faces to train to recognize faces without mask and using faces without mask to train to detect masked faces.
3. The AFR-Conv algorithm integrated into the ConvMixer model is a novel approach for handling face occlusion. By assigning priority-based weights to different face patches and using residual connections, the algorithm can effectively focus on relevant facial regions, even when faces are partially occluded, leading to improved recognition accuracy in challenging real-world scenarios.
4. The introduction of the ConvMixer architecture specifically tailored for face recognition tasks is a significant contribution. ConvMixer's ability to capture complex spatial patterns in face images efficiently makes it a powerful feature extractor, enhancing the model's discrimination and recognition capabilities.
5. The ConvMixer and AdaBoost approach offers lightweight and efficient feature descriptors. This characteristic is vital for real-time processing in video surveillance systems, where computational complexity is a significant concern.
6. The experimental results demonstrate that the proposed ConvMixer and AdaBoost-based face recognition system outperforms advanced methods for face classification. This superiority showcases the system's competitiveness and effectiveness compared to other existing deep learning algorithms.

1.3. Paper Organization

The remainder of the paper is structured as follows: Section 2 presents a recent survey of past studies in the field of occluded face recognition, especially using DL techniques. Section 3 demonstrates the data acquisition process and the proposed methodology. Section 4 presents the experimental results and comparisons to other techniques.

Section 5 discusses the results attained. Finally, Section 6 summarizes the main conclusions of this paper.

2. Literature Review

For law enforcement, FR is an appealing area of research and development. Surveillance cameras are used in conjunction with intelligence techniques worldwide to detect criminal activity. Currently, as the epidemic of COVID-19 spreads all over the world, people are forced to cover most of their faces with masks to contain the pandemic, requiring much more accurate face recognition algorithms for identity verification. Factors associated with biometric sample capture and presentation, such as facial occlusions, have a significant impact on the precision of FR algorithms [10], [11].

Past studies showed different problems existed in recognizing human faces in real-time: 1) Face pose: Computerized systems are highly sensitive to pose variations. When a person's head and viewing angle vary, so does his or her facial position. 2) Illumination condition: The variation of lighting conditions has a significant impact on the quality of an image. 3) Face occlusion: The biggest challenge for computer vision systems is recognizing human faces when they cover their faces with masks. 4) Expressions: Varied conditions cause multiple human moods, which lead to the display of various emotions and, subsequently, changes in facial expressions. 5) Aging: The appearance of a person's face varies over time and reflects their age, which is a new problem for facial recognition algorithms. Many researchers presented techniques for occluded face recognition. [12] The authors developed an automatic facial recognition solution. The settings involved masked probes, unmasked pairs, masked pairs, and unmasked references with actual and synthetic masks.

In [13], the author developed an end-to-end FR network that is insensitive to face masks and invariant to face images. First, face mask synthesized datasets were created by accurately matching the face mask to images in publicly available datasets, namely LFW, CASIA-Web Face, CFP, CPLFW, and CALFW. Afterward, datasets were used to generate training and testing datasets. Second, they introduced a model consisting of two components: an alignment component and a feature extraction component using DCNN to generate a 512-feature vector. The network is invariant to face images with a face mask since these modules are trained end-to-end. Their experimental work showed significant improvement compared to state-of-the-art systems. The authors of [14] proposed a CNN model for face detection based on facial features. They developed a new method for detecting faces based on the spatial structure and arrangement of facial components' responses. The grading system is data-driven, and it was carefully crafted to account for difficult circumstances where faces are only partially visible. Faces with extreme occlusion and unrestricted pose fluctuations are detected by their CNN architecture. On well-known benchmarks, namely, AFW, PASCAL Faces, WIDER FACE, and FDDB, their technique performs admirably.

In [15], the authors proposed a set of repurposed datasets as well as a standard for researchers to employ. They also presented a pre-training method based on visual representation learning tailored to unmasked vs. masked face matching. Their research discovered robust traits that might be used to distinguish people in a variety of data collection circumstances. This is accomplished by training on a variety of datasets and confirming our results using a variety of holdout datasets. When it came to masked-to-unmasked face matching, their method's specific weights outperformed conventional face recognition features. The authors introduced a mask-aware FR system in [16] that can distinguish between people wearing and not wearing facial masks. They evaluated three traditional descriptors, such as local binary pattern (LBP), local directional order pattern (LDOP), and histogram of oriented gradients (HOG), along with support vector machine (SVM) for face mask recognition. In addition, they created a mask-aware dynamic model based on deep learning that can distinguish faces in the

presence and absence of facial masks. The real-world masked face recognition dataset was used in the evaluation. LDOP-based descriptors had the maximum accuracy of 99.60% in facial mask detection. In the presence of a facial mask, their proposed dynamic ensemble model has 99.53% accuracy.

Table 1 Comparison of affective states-related work.

Cited	Description	Techniques	Dataset	Results
[13]	End-to-end FR network that is not directly impacted by face masking	DCNN	CASIA, Masked LFW, CALFW, CPLFW, Masked CFP-FF	ACC:75.50% (CASIA) 98.41% (LFW) 86.15% (CALFW) 79.42% (CPLFW) 94.44% (CFP-FF)
[14]	DL model for face detection under severe occlusion and unconstrained pose variations	CNN	FDDB, PASCAL Faces, AFW, and WIDER FACE	Recall: 92.84% (for FDDB)
[16]	proposed a mask-aware face recognition system	SVM ResNet-50	RMFRD	ACC: 99.53%
[17]	A face mask detection model that combines deep and traditional machine learning.	ResNet-50 SVM	RMFD SMFD LFW	ACC: 99.64% 99.49% 100 %
[18]	A entire training framework for ArcFace-based facial recognition models, allowing them to be adapted to function with masked faces.	LResNet-50	MS1MV2 Masked LFW Masked CFP-FF Masked CFP-FP	ACC: 99.78% 98.92% 98.33% 88.43%
[19]	The Additive Angular Margin Loss function can improve the discriminative power of feature embeddings learned with DCNNs for FR.	ResNet-100	IJB-B LFW CALFW CPLFW	ACC: 94.2% 99.82% 95.45% 92.08%

In [17], a hybrid face mask detection model was proposed that combined deep and traditional machine learning. There were two phases to the proposed framework. The first component was created to extract features using Resnet 50. The second component was created to help with the classification of face masks utilizing SVM, decision trees, and an ensemble approach. The investigation focused on three face-masked datasets. The Simulated Masked Face Dataset (SMFD), Labeled Faces in the Wild (LFW), and RMFD are the three datasets that were used, and an accuracy of 99.49%, 100%, and 99.64% were achieved, respectively, on the test datasets. The authors proposed a complete training pipeline based on the loss function [18] and ArcFace model [19], with numerous changes to the backbone and loss function. They used a ResNet-50 model as a backbone. For MS1MV2, they achieved a mask-usage detection accuracy of 99.78%. They presented experimental results for 10 different face recognition benchmarks. Their findings showed that their strategy regularly exceeded the state of the-art in extensive tests.

The COVID-19 outbreak led to masked face recognition (MFR) development [20], but overemphasizing it harms standard face recognition. MFR should be treated as a mask bias, not a separate task. The study examined how face masks influenced emotion recognition in first- and fifth-graders, along with young adults [21], considering mask

presence, color, and emotion type. Results showed masks affected recognizing fear and sadness, but not anger or neutrality. This study [22] aims to create an attendance system using face recognition and mask detection, accessible online via a browser interface. No special software installation is needed; users can access it through any terminal. The system records attendance data centrally in an online database, utilizing biometric face signatures. Users' profiles are loaded with face-image samples. Initial steps involve SVM-based model training for face recognition and synthetic data for identifying masked users. The goal is an efficient system for attendance management, even with face masks. In response to widespread mask-wearing during COVID-19 [23], conventional face recognition struggles. This article proposes an eyebrow-focused network for masked face recognition, using local features like eyebrows due to limited visible cues. The approach includes feature extraction, eyebrow pooling, and fusion using a graph convolutional network. Tested on real-world and synthetic datasets, the method outperforms existing techniques, effectively addressing masked face recognition challenges.

DeepMasknet [24] was introduced to deal with mask-wearing issue. They also created a new diverse dataset, MDMFR, for evaluation. DeepMasknet outperforms existing models across datasets, providing a solution for COVID-19 challenges. COVID-19 challenges traditional face recognition due to increased mask-wearing [25]. Limited facial data hampers recognition, prompting experiments with CNN architectures and altered methods. The study evaluates existing CNN-based systems using entirely masked face datasets, showing the importance of network depth and suggesting adjusted parameters. Empirical analysis guides new parameter values for masked face recognition.

The paper introduces a method to improve face recognition with masks [26]. It employs mask transfer for data augmentation and presents Attention-Aware Masked Face Recognition (AMaskNet) consisting of a feature extractor and a contribution estimator. Amid COVID-19, mandatory mask use prompted the development [27] of a system recognizing people wearing masks from photos. Using MobileNetV2 and OpenCv's face detector, the model detects faces and identifies mask presence. FaceNet extracts features, and a multilayer perceptron performs recognition. Training on 13,359 images (52.9% masked, 47.1% unmasked), the system achieves 99.65% accuracy in mask detection, 99.52% in recognizing masked individuals, and 99.96% for unmasked recognition. The research addresses mask-related challenges in facial recognition, yielding high accuracy in both mask detection and recognition tasks.

An improved solution [28] for masked face recognition is proposed, merging a cropping-based method with the convolutional block attention module (CBAM). The approach optimizes cropping and employs CBAM to emphasize eye regions. Unique scenarios of using unmasked faces to train for masked recognition and vice versa are explored. Extensive experiments on various datasets demonstrate the approach's superiority over other methods, notably enhancing masked face recognition performance. This article introduces [29] a robust face recognition method called FROM (Face Recognition with Occlusion Masks) to handle occlusions. It employs a single end-to-end deep neural network to identify and correct corrupted features using dynamically learned masks. A vast dataset of occluded face images is used for effective training. Unlike other methods relying on external detectors or shallow models, FROM is both simple and powerful. Experiments on various datasets confirm that FROM significantly enhances accuracy under occlusions and performs well in general face recognition scenarios. In response to the global need, this paper offers a straightforward solution [30] using TensorFlow, Keras, OpenCV, and Scikit-Learn for face mask detection. The approach efficiently identifies faces in images/videos and determines mask presence. It handles faces with masks in motion and videos for surveillance purposes, achieving high accuracy. The study fine-tunes optimal parameters for Convolutional Neural Network (CNN) models to accurately detect masks without overfitting. In the presence of Covid-19 masks, Table 1

compares the existing approaches for detecting and recognizing faces in obstructed environments.

3. Materials and Methods

We split the RGB image into non-overlapping patches of size 4×4×3 = 48 on the transformer layer in the first stage. A linear embedding layer is used in the encoder's initial stage to modify the patch tokens' feature dimension. There are NA shifting window attention blocks (Fig. 3a) inside each transformer layer, which have linear computing complexity and cross-window connections to deal well-separated localities, creating the architecture for modeling useful image-level attributes. We compress our feature maps from H 4 × W 4 to H 8 × W 8 for a hierarchical representation by patch fusing layers for the subsequent phase. We also add a relative position embedding (RPE) in [3, 22, 23, 37, 40], where RPE ∈ RN × N and N = M × M are the sequence lengths and M is the window size.

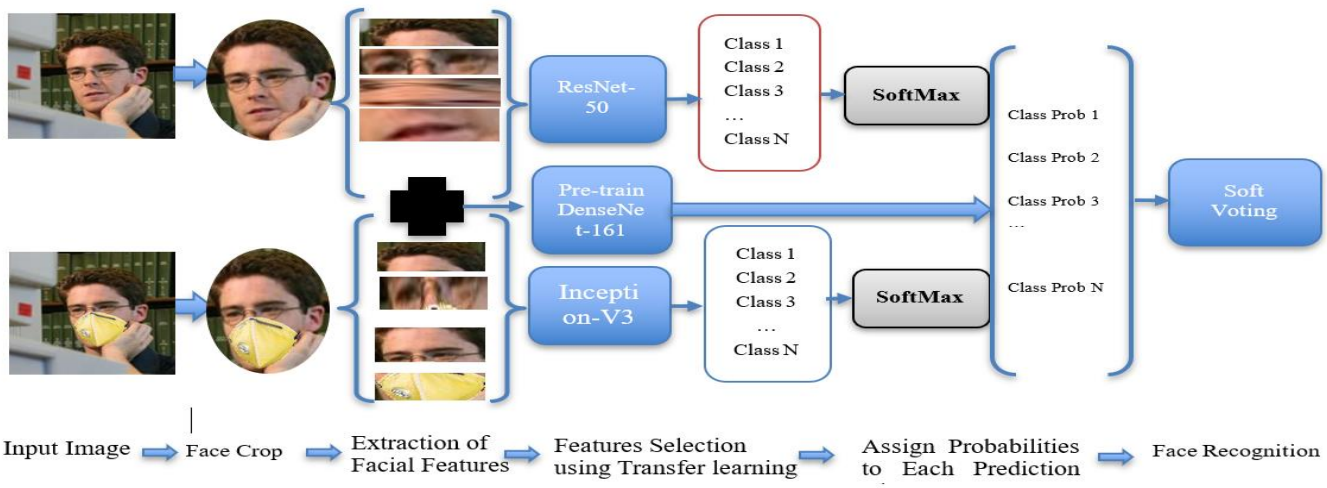

Figure 1. Flow Chart of Proposed deep learning face recognition architecture based on facial masks

Automated face recognition (AFR) with occlusion is a challenging task. Therefore, a ConvMixer [33] model uses image patches and convolutional operations to extract compelling features. The ConvMixer model shows superiority compared to Vision Transformers due to image patches. They use depthwise convolution followed by pointwise convolution as their main ConvMixer layer, which is repeated depth times. They use the Gaussian error linear unit (GELU) activation function. It gives better performance with a large kernel size. It provides better validation accuracy than a basic CNN model with four times fewer parameters.

Even though AFR identification has been used to replace previous approaches, it is still susceptible to occlusion. A new FAR method based on residual connections and a convolutional mixer (Conv-Mixer) network with AdaBoost to recognize human faces. The Conv-Mixer and a residual connection-based CNN model are utilized first to extract features from facial images. The photographs are then classified using the AdaBoost machine-learning classifier. This work employed a skip connection to maintain the model's performance or prevent gradient degradation. The AFR-Conv deep-learning model makes use of three residual connections. To reduce computational complexity, the Conv-Mixer is applied instead of traditional convolution. The idea behind using the Conv-Mixer with a residual connection-based CNN model is to develop improved representations of input data while keeping the model's computational cost low. The pro-

posed CNN model's features map is given to the AdaBoost classifier to recognize a human's identity. Instead of a fully connected layer, an AdaBoost classifier is used since it is straightforward to use and is thought to be helpful in classification jobs to improve performance. The flowchart of a typical AFR approach is shown in Figure 3. The significant steps of the AFR-Conv model created in this study are described in the following subsections.

The architecture developed in this paper is based on a novel AFR-Conv (Automatic Face Recognition Convolutional) model for human face recognition. The key components and steps involved in the architecture are visually represented in Fig. 1 and explained in the subsequent paragraphs.

The model takes an image of a human face as input. The architecture begins with two convolutional layers, each using a kernel size of 3. Convolutional layers are essential for feature extraction in convolutional neural networks (CNNs). After each convolutional layer, there is a max-pooling layer. Max-Pooling reduces the spatial dimensions of the feature map while retaining the most essential information. There are eight batch normalization layers in this architecture. BN is applied before the activation function after every separable convolution layer. BN normalizes the activations, leading to faster and more stable training of deep neural networks. The architecture employs three residual blocks, each containing two separable convolutional layers. Separable convolutions are depthwise convolution followed by pointwise convolution, which reduces computational complexity and enhances performance. After the three residual blocks, there is one final convolutional layer with a kernel size of 1. This layer is used to refine the features further. All the above layers together produce a feature map, a condensed representation of the input image that encodes essential facial features. The feature map is flattened to convert it into a 1-dimensional vector, ready for further processing. Following the flattened layer, there is one dense layer. Dense layers are fully connected layers that perform classification based on the learned features.

AdaBoost Classifier with Linear Activation: AdaBoost is a machine learning algorithm that combines multiple weak classifiers to form a robust classifier. This architecture uses it as the final classifier for face recognition. The weak classifiers could be the combination of previous layers' outputs, and their outputs are linearly combined to make the final prediction. The GELU activation function calculates kernel weights' output value in every convolutional layer and the last convolutional layer. GELU is known to enhance classification accuracy.

It is important to note that this description provides an overview of the proposed architecture but needs more specific details such as the number of filters, the size of the fully connected layer, and the number of boosting iterations in AdaBoost. The architecture is tailored for human face recognition, utilizing multiple layers for feature extraction and a boosting-based classifier for final decision-making. However, the model's effectiveness would require empirical evaluation and comparison with other state-of-the-art face recognition models on suitable datasets.

3.1 Data Acquition

The data is gathered from a variety of popular datasets available on the Internet, as described in Table 2. Faces with masks appear in a small number of data sets. As a result, an augmentation approach is used on multiple common verification datasets to create the face mask synthesized evaluation dataset. The data augmentation technique is applied to LFW [31], CALFW [32], CPLFW [33], and CFP [34]. The LFW (Labeled Faces in the Wild) is a popular public face verification benchmark containing 13K photos and 5.7K IDs. To analyze the performance of the suggested AFR-Conv, 8500 face photos with masks were employed in total. Cross-Age LFW (CALFW) is a revision of LFW that stresses the age disparity between positive couples even more to increase intra-class variation. CPLFW (cross-pose LFW) is a revision of LFW that stresses pose differences to increase intra-class

variation. Frontal-Profile (CFP) is a FR dataset created to aid studies into the challenge of in-the-wild frontal-to-profile face verification. The CFP's frontal-profile and frontal-frontal verification pairings are employed in this paper. Only frontal face pictures are synthesized using face masks due to the high percentage of unsuccessful landmark detection in profile photographs. Figure 2 (a) illustrates an example of the LFW dataset's generated face mask-enhanced pictures.

To avoid overfitting, data augmentation is used. To increase the variance in the training dataset, the data is augmented by mirroring and cropping the photos. Each preprocessed face picture in the training set is supplemented into four images after the preprocessing stage by rotating the input image in four directions: 0°, 90°, 180°, and 270°. Augmentation aids in boosting data size, producing new data from existing data, and overcoming the absence of labeled pictures. The Real-World Masked Face Dataset (RMFD) [35] is used in addition to the synthetic versions of typical face recognition testing datasets. The world's biggest masked face dataset, according to RMFD, is the world's largest masked face dataset at the time of writing. From cleaned and annotated photos scraped from the internet, the dataset comprises 5,000 masked faces of 525 individuals and 90,000 normal faces. Figure 2(b) depicts photos from the RMFD dataset with and without a face mask.

Table 2: Different public face datasets statistics and selected images for experiments

Dataset	#Images	#Identities	Select Images	Web link
CASIA-Webface [30]	494,414	10,575	600	https://paperswithcode.com/dataset/casia-webface
LFW [31]	13,233	5,749	2000	http://vis-www.cs.umass.edu/lfw/
CALFW [32]	12174	4,025	1000	http://whdeng.cn/CALFW/index.html
CPLFW [33]	12174	4025	3000	http://www.whdeng.cn/cplfw/index.html
CFP [34]	10 per identity 4 profiles / identity	500	400	http://www.cfpw.io/cfp-dataset.zip
RMFD [35]	5000 with mask 90,000 without mask	525	500	https://github.com/X-zhangyang/Real-World-Masked-Face-Dataset
Proposed FR-OIF			8600	

3.2 Extract Facial Features

The first and most important step in an automated face recognition system is face detection. A face image is used as an input in the face detection method, and the output is used to detect the exact individual from the dataset.

Face feature extraction extracts geometrically formed facial features for face identification. [36]. First, eye detection considers the face map, which is the output of face detection and cropping. Face edges are recognized after the face mapping process. Gabor filters are used to create a filtered face image. Gabor kernels in two dimensions are used. The generic eye detector is given a filtered face image and uses a Fast Transfer Learning method based on support vector machines to detect the eye appearance from other facial features (MultiFTLSVM) [37]. The MultiFTLSVM classifier's fundamental idea is to create a hyperplane that isolates eye features from other features. It obtains eye, nose, and

mouth sub-images based on geometrical considerations and extracts the fiducial points from the detected eye centers.

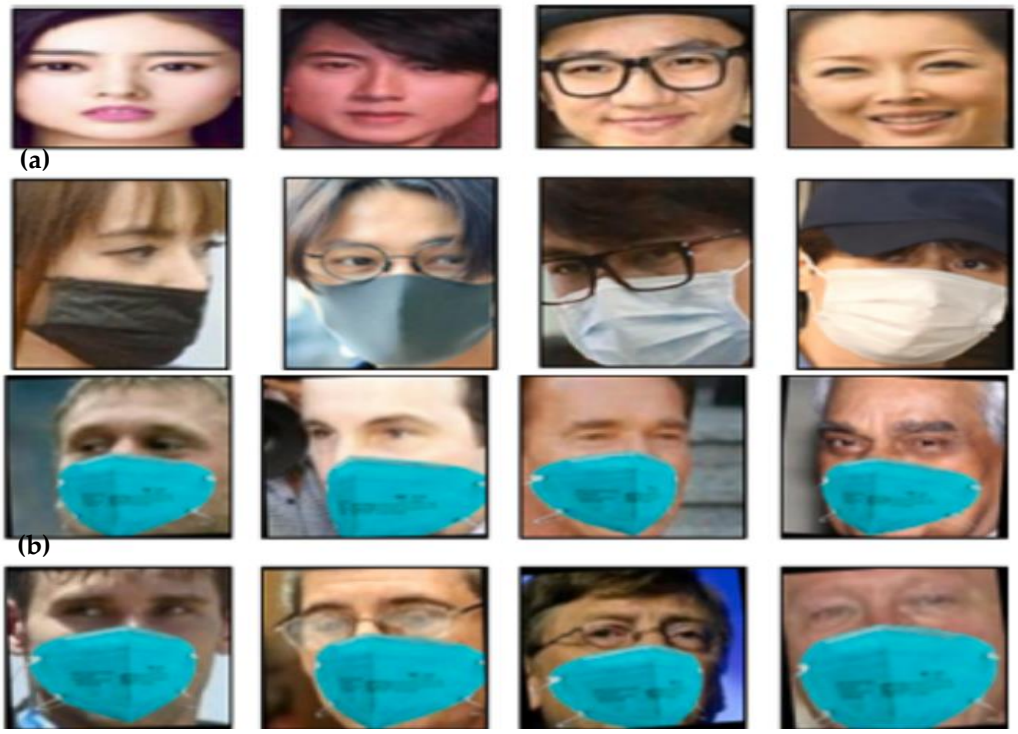

Figure 2: Sample of the original face mask with synthesized and figure (a) shows the LFW face dataset and figure (b) represents the sample of the RMFD face dataset.

3.3 Pretrained Transfer Learning

The proposed face recognition system employs three of the most powerful pre-trained CNN models, which are: ResNet-50 [38], Inception-v3 [39], and DenseNet-161 [40]. There are also several disadvantages related to CNN models. The two most significant drawbacks are the lengthy processing period and the over-fitting issue. Because of the processing time needed, a deep learning model [41] is difficult to implement on a single normal computer system with few CPUs. Fortunately, graphics processing units (GPU) have solved this problem as technology has advanced [42]. The deep learning model can be used in real-world settings by combining numerous CPUs and GPUs. There is also an issue of overfitting with the CNN model. They are trained on millions of learnable parameters, as previously stated. Therefore, CNN-based systems usually require a large amount of training data. Although numerous strategies have been employed to reduce this issue, such as data augmentation and dropout, the amount of training data in such CNN systems remains enormous. To deal with the problem recently, the transfer learning method was adopted [35, 43]. The transfer learning method allows us to apply a CNN that has been trained with enough training data for one problem to another. This strategy has been found to be useful in several situations, especially when significant amounts of training data are sparse, such as in medical imaging [44] or finger vein recognition [45].

Figure 3 shows a comparison of the transfer learning scheme to the conventional ML method. As shown in this figure, the transfer learning approach learns system information from two sources: the challenge to be solved ("target task") and knowledge (a model) gained from a previous machine learning problem. In a traditional machine learning system, the system model is only learned for a single job using a single source of data. CNN can be reused and transferred to a new problem using the transfer learning method. We modified the DenseNet-161 model for our experimental work and used it to

construct the proposed CNN architecture. The Image-Net dataset was used to pre-train the VGG16 model. Section 3.3 describes the design of the proposed model FAR-Conv. Furthermore, the fully connected layer was employed as the last layer for classification in the pre-trained DenseNet-161 model. The AdaBoost classifier is used to distinguish human faces with occlusions in the proposed improved model.

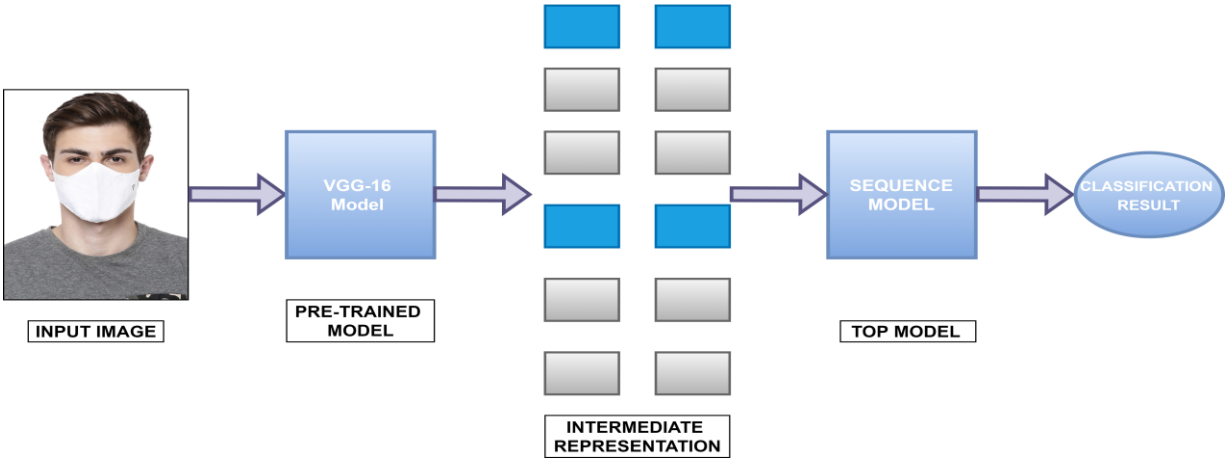

Figure 3. Proposed deep learning face recognition architecture.

3.4 Proposed ConvMixer Learning Model

The AFR-Conv system is developed in this paper based on the trained ResNet-50 model. The AFR-Conv system architecture is depicted in Figure 2. For object recognition, a novel architecture based on depth-wise separable convolutions was recently proposed. Tolstikhin et al.'s MLP-Mixer model [44, 45] was used as inspiration for architecture. To be more specific, a depthwise convolution is used to mix spatial locations before a pointwise convolution is used to combine channel locations. Figure 3 shows the ConvMixer blocks' modified version of the original ConvMixer layer. The batch normalization operation and activation layers are switched in order from the original version. We also utilize ReLU instead of GELU to activate all layers. The DSC offers two advantages when it comes to constructing a deep learning model: 1) It might be able to reduce the number of parameters, and 2) It may be used to improve model generalization. Thus, DSC was found to improve training efficiency and classification accuracy.

The ConvMixer architecture tries to prove that the superiority of the ViT is partly due to using image patches and introduces a novel ConvMixer model that is like the ViT as well as the MLP-Mixer model. It works directly with patches as input, isolates the mixing of spatial and channel dimensions, and keeps the network's size and resolution constant. But it utilizes convolutions to achieve the mixing steps. It gives better validation accuracy compared to a basic CNN model with four times fewer parameters. It also uses batch normalizations instead of layer normalizations.

The ConvMixer model is a recent approach that highlights the power of processing images in patches to achieve impressive performance on various tasks. Its architecture consists of first splitting input images, each of 32x32 pixels with three RGB channels, into different patches, enabling local information processing. The crux of ConvMixer lies in the alternating application of convolutional networks along the channel-wise and space-wise dimensions of these patches. This approach allows the model to capture cross-channel interactions and local spatial relationships effectively. Without the need for recurrent layers or self-attention mechanisms, ConvMixer demonstrates remarkable results by assembling basic building blocks like convolutions, nonlinearities, batch normalizations, mean pooling, and dense layers in different architectures. This simple yet

potent model sheds light on the significance of patch representations for high-performance image understanding and classification tasks. Further insights and specific architectural details can be found in the original ConvMixer paper. The architecture of ConvMixer is summarized in Figure 4.

The main concept of the ConvMixer architecture is to begin by splitting the input image into patches of size (p, p) using a convolutional layer with the stride argument. The stride determines how the convolutional kernel moves across the input image. If the stride is set to 1, the convolutional kernel is applied around every pixel in the image, resulting in overlapping patches. The overlap occurs because the kernel moves one pixel at a time, covering neighboring regions. On the other hand, if the stride is set to a value greater than 1 (e.g., stride = 2), the convolutional kernel skips pixels, only applying the convolution to every other pixel. As a result, the patches become non-overlapping and cover the image in a grid-like fashion. When stride = p, the convolutional kernel moves p pixels at a time, leading to disjoint and adjacent windows. These windows cover the entire image in non-overlapping patches of size (p, p). Each patch is then processed independently through the ConvMixer architecture, allowing the model to focus on local information and efficiently capture spatial relationships within each patch.

This patch-based processing is a fundamental aspect of ConvMixer's design, enabling the model to capture fine-grained features and achieve impressive performance on various tasks without the need for complex recurrent or attention mechanisms. Therefore, the first layer of ConvMixer is:

$$Z_0 = BNorm\big(\sigma Convol \to h(X, stride = p, kernelsize = p)\big) \quad (1)$$

The second part of the model is the main ConvMixer layer which is repeated depth times. This layer consists of residual block containing a depthwise convolution. A residual block is nothing but a block where the output of a previous layer is added to the output of another later layer. In this case the inputs are concatenated to the output of the Depthwise convolution layer. This output is followed by the activation block which is then followed by a pointwise convolution and another activation block.

$$Z_l = BNorm(\sigma ConvolDepthwise(Z_{l-1})) + Z_{l-1} \quad (2)$$

and

$$Z_{l+1} = BNorm(\sigma Convolpointwise(Z_l)) \quad (3)$$

The third part of the ConvMixer model involves a global pooling layer to obtain a feature vector of size h from the processed patches. Global pooling reduces the spatial dimensions of each patch to a fixed size, which can then be passed to a SoftMax classifier, depending on the specific task. The activation function used in ConvMixer is GELU (Gaussian Error Linear Unit). GELU is a smooth and differentiable activation function that is known to perform well in deep neural networks. Unlike ReLU (Rectified Linear Unit), which sets all negative values to zero, GELU weighs the inputs based on their magnitude rather than gating them based on their sign. This characteristic of GELU allows it to preserve both positive and negative information in the activation, making it suitable for models like ConvMixer.

$$GELU(x) = x.\varphi(x) \quad (4)$$

This smooth non-linearity helps in reducing the issues of "dying ReLU" where neurons get stuck and stop learning due to being always inactive (zero gradient). Overall, the global pooling and GELU activation contribute to the final feature representation of the image patches, enabling the ConvMixer model to produce a compact and informative feature vector that can be used for downstream tasks such as image classification or object detection.

The patch embedding in the ConvMixer model summarizes a p × p patch from the input image into an embedded vector of dimensions e. The embedding process is achieved through a single convolutional layer with a kernel size of p, a stride of p, and h output channels. This convolutional operation takes the p × p patch as input and trans-

forms it into a new representation with h channels. The result of the convolutional operation is then passed through a non-linearity, which introduces non-linearity to the embedding process. The non-linearity can be the GELU activation function, which has been previously mentioned as the activation function used throughout the ConvMixer model.

This patch embedding trick is used to convert the entire n × n image into a feature map with dimensions h × n/p × n/p. Each h × n/p × n/p feature map corresponds to the embedded representation of a particular patch of size p × p. To normalize the output of each layer and stabilize the training process, batch normalization is applied after each convolutional layer in the ConvMixer model. Batch normalization centers and scales the activations within a batch along each dimension, introducing learnable parameters for the mean and standard deviation. In this framework, BatchNorm(H) is used to apply batch normalization after the convolutional layer, where H represents the number of output channels from the convolution operation.

By incorporating patch embedding and batch normalization, the ConvMixer model can effectively process patches of the input image and extract meaningful features, enabling it to achieve remarkable performance on various tasks.

Algorithm 1. Proposed Automatic Face Recognition Using ConvMixer CNN

Input:	Read Tensor X
Output:	Feature map Extracted x = (x1, x2,x3,…,xn)
Step 1.	**Data Augmentation and Preprocessing**
Step 2.	To begin, create the essential functions (a) Conv-Batch Norm and (b) Separable ConvBatch Norm.
Step 3.	The Conv-Batch Norm block accepts tensor X, which contains a number of filters, and kernel size as inputs. a) X is given a Convolution layer. b) After that, Batch Normalization is used.
Step 4.	We utilized Separable Conv2D instead of Conv2D in the Conv-Batch Norm Block in Step 2.
Step 5.	**Model Construction**
Step 6.	a) There are two Conv layers with 32 and 64 filters each. A ReLU activation follows each of these. b) Then, using Add, Skip Connection is used. c) There were three skip connections. Two Separable Conv layers precede MaxPooling in each Skip Connection. The skip connection has Conv of 1 × 1 with strides 2.
Step 7.	After that, the feature map x = (x1, x2,...., xn) was created and flattened using the flatten layer.

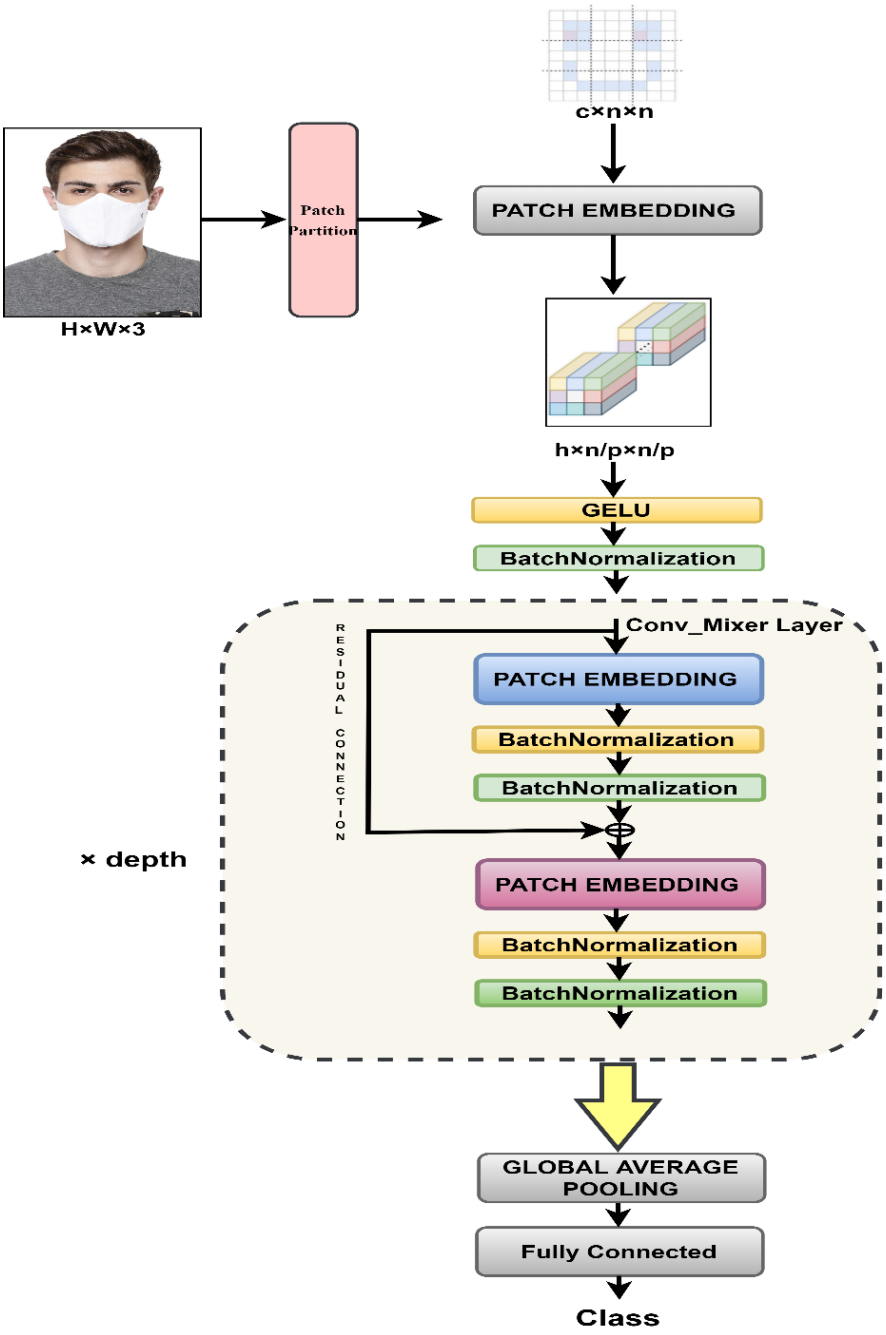

Figure 4. The proposed automated face recognition (AFR) method is based on a residual connection with ConvMixer to extract features and classify them by AdaBoost.

3.5 Deep Residual Network Connections

The terms "residual connections" and "skip connections" are interchangeable. They are utilized to allow gradients to flow directly through a network, bypassing non-linear activation functions. The non-linear character of non-linear activation functions causes gradients to erupt or vanish (depending on the weights). Skip connections resemble a 'bus' that travels the length of the network, with gradients flowing backwards.

The residual link, also known as a skipped connection, skips the two or three tiers of the network. Figure 5 depicts the DL network's solitary remaining connection block. As shown in Figure 1, there are three residual blocks in our proposed CNN model. The

benefit of using residual connectivity in a DL model is that the function from the previous layer is added to the next layer by the preceding levels of the model network. A shortcut link, as shown in Fig. 5, defines the residual network by transforming the network building block into its residual counterpart. The identity mapping shortcuts mentioned in Eq. (5) can be used directly when the input and output dimensions are the same.

$$y = F(x, \{W_i\}) + x \quad (5)$$

The building block is changed to a bottleneck building block for computational reasons. Instead of two layers, a stack of three layers is employed for each residual function F, as shown in Fig. 5. The three layers are 1×1, 3×3, and 1×1 convolutions, with the 1×1 layer lowering and then raising (restoring) dimensions, and the 3×3 layer acting as a bottleneck with reduced input/output dimensions. Practical concerns have led to the use of the bottleneck building block. Furthermore, the bottleneck construction block is caused by the deterioration problem of plain networks. The architectural layers of ResNet-50 are depicted in Table 3.

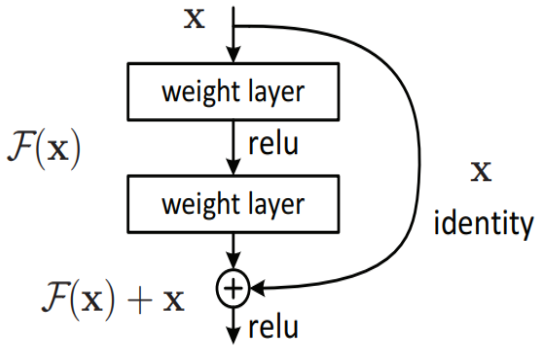

Figure 5: ResNet-50 Bottleneck building block

The genuine output value is H(x), and the residual learning of layers in the network input x is F(x).

Table 3: ResNet-50 architectural view

Layer Name	Layer Type	Input Size	Output Size
Input Image	Input	(32, 32, 32, B)	(32, 32, 32, B)
Patch Split	Convolution (Stride = p, Kernel = p)	(32, 32, 32, B)	(h, n/p, n/p, B)
ConvMixer Block 1	Alternating Convolutional Layers	(h, n/p, n/p, B)	(h, n/p, n/p, B)
Skip Connection 1	Elementwise Addition	Same	Same
ConvMixer Block 2	Alternating Convolutional Layers	Same	Same
Skip Connection 2	Elementwise Addition	Same	Same
ConvMixer Block 3	Alternating Convolutional Layers	Same	Same
Skip Connection 3	Elementwise Addition	Same	Same
Global Pooling	Global Average Pooling	Same	Same
Flatten	Flatten	(h, 1, 1, B)	(h * B,)
Dense Layer	Dense	(h * B,)	(e, B)
SoftMax	Softmax	(e, B)	(num_classes, B)

The input size (32, 32, 32, B) represents the initial image size with B being the batch size, h is the number of output channels from the patch embedding layer, and n/p is the resulting spatial dimension after patch splitting. The ConvMixer blocks consist of alternating convolutional layers applied channel-wise and space-wise. Skip connections are

added after each ConvMixer block to directly add the output of the block to its input. This helps avoiding the vanishing gradient problem and allows the model to go deeper effectively.

Global average pooling is applied to get a global representation of the feature map.

The final Dense layer is used for classification, and the SoftMax activation function is applied to produce the probabilities for each class. The output size is (num_classes, B), where num_classes is the number of classes in the classification task. Please note that the actual values of h, p, e, and num_classes depend on the specific configuration and requirements of the ConvMixer model, and the ResNet-50 architecture being used. The table provides a general outline of how skip connections can be incorporated into the ConvMixer model to make it deeper and more powerful, similar to ResNet-50.

3.5 Features Classified using AdaBoost Classifier

The Adaboost [46] algorithm is an ML technique for FR that uses eigenvalues to extract features. The AdaBoost algorithm is used to develop a powerful learner over several rounds. AdaBoost creates a powerful learner by layering weak learners on top of one another. A new weak learner is added together, and a weighting vector is adjusted to focus on examples that were misclassified in previous rounds to produce a strong classifier employing numerous classifiers while training the data set. Face recognition analysis is widely employed in a variety of applications. According to a review of the literature, various algorithms have been created to recognize faces. The AdaBoost method is a simple-to-implement algorithm that improves detection accuracy. As a result, this research evaluates an AdaBoost algorithm for human face recognition.

All samples are equally weighted with Wi during the AdaBoost training phase. The weights are then repeatedly improved by raising the weights associated with misclassified data. To generate the final output of the boosted classifier, numerous weak learners can be combined in a weighted sum using the AdaBoost process. When compared to other commonly used classifiers such as neural networks and SVM, AdaBoost may achieve good classification performance with fewer parameter adjustments. We only choose a weak classifier for the specified classification problem; and (ii) the number of boosting steps used in the training step when implementing AdaBoost. Each round of boosting can include many weak classifiers. At that round of boosting, the AdaBoost algorithm will choose the weakest classifier that gives the best results. The following are the major processes involved in implementing the AdaBoost algorithm: To implement it, we need decision stumps, which work on the principle of the AdaBoost classifier. The procedure is carried out three times. A linear combination of weak classifiers makes up the final classifier.

AdaBoost must meet two requirements: (1) the classifier must be trained interactively on a variety of weighed training instances; and (2) the classifier must be trained on many weighed training examples. It seeks to minimize training errors in each iteration to produce a good fit for these samples. What is the mechanism behind the AdaBoost algorithm? The procedure is as follows: AdaBoost begins by randomly selecting a training subset. It trains the AdaBoost machine learning model iteratively by picking the training set based on the previous training's accurate prediction. It gives incorrectly categorized observations a larger weight so that they have a higher chance of being classified correctly in the next iteration. Eq. (6) represents this state as follows:

$$S(x) = w_1 s_1(x) + w_2 s_2(x) + w_3 s_3(x) + w_4 s_4(x) \quad (6)$$

where S is a string classifier, w is the weight parameter and s1, s2 are weak classifier and x is a feature vector in Eq. (6). The sign of s1(x) decides to which class point x is assigned, by the ith weak classifier and sign of s(x) decides to which class point x is assigned by the final strong classifier. In addition, it distributes weight to the trained classifier in each iteration based on the classifier's accuracy. The classifier with the highest accuracy will be given the most weight. This process is repeated until all the training

examples fits nicely, or the largest number of predictions has been reached. Perform a "vote" across all the learning algorithms you created to categorize them.

Algorithm 2. AdaBoost Classifier to Recognize Human Faces

Input:	Input Extracted Feature map x = (x1, x2,x3,…,xn) with labels Y.
Output:	Class Labels, Y=1,0 where 1 shows the recognize and non-recognize face and test data xtest.
Initialize:	Weights w1;i= 1/ 2l or1/2m for yi or, respectively, with l+m=n. where m and l are positive and negative samples.
Process:	
Step 1:	Construction of AdaBoost Classifier for Recognizing human faces (a) The AdaBoost classifier is trained using feature samples x= (x1,x2,x3,...,xn) derived from the proposed ConvMixer deep learning architecture, which includes both positive and negative data. (b) Use Eq. 4,5 and 6 to generate week classifiers and update weight over misclassified samples.
Step 2.	Combine the week classifier to generate strong classifier for recognize human identity.
Step 3.	The decision function of the below equation is used to allocate test samples xtest to a class label. Xtest = (w.x)+b.

3.6 Fine-tuned Model and Hyperparameters

In this face recognition example, we begin by preparing a dataset containing face images, which we split into training, validation, and test sets. For the ConvMixer architecture, we adopt a simplified version consisting of a single layer with a convolutional step, followed by LayerNorm, ReLU activation, and a Feedforward Mixer with ReLU activation. The model's weights are initialized using He initialization. During training, we employ a fixed learning rate of 0.001 and perform data augmentation with a batch size of 32 and a dropout rate of 0.2 to regularize the model. The goal is to minimize the cross-entropy loss function, as it's well-suited for classification tasks like face recognition.

Next, we integrate an AdaBoost classifier into the system. The ConvMixer model acts as the base classifier, and we train AdaBoost with 50 weak learners and a learning rate of 0.1. The AdaBoost algorithm will combine the outputs of these weak learners to form a strong classifier, enhancing the overall performance of face recognition.

Throughout the process, we conduct a hyperparameter search to fine-tune the model effectively. This involves experimenting with various hyperparameter combinations to optimize the ConvMixer's performance on the validation set. In cases of overfitting, we consider implementing early stopping to prevent excessive training. Finally, we evaluate the fully trained AdaBoost classifier on the test set to get an unbiased estimate of its performance in recognizing human faces. By iteratively adjusting the model architecture and hyperparameters, we aim to achieve the best possible accuracy in face recognition, making this approach applicable to real-world scenarios involving video surveillance and urban security.

3.7 System Implementation

The AFR-Conv system outlines an approach for automated face recognition that combines multiple advanced techniques to achieve accurate results. These steps are described in algorithm 3. It begins by initializing parameters, including the number of ConvMixer blocks and AdaBoost iterations, and selecting powerful pre-trained CNN models such as ResNet-50, Inception-v3, and DenseNet-161. The preprocessing step prepares the training and testing images for analysis. The algorithm then sets up the ConvMixer architecture, including ConvMixer blocks and skip connections. AdaBoost is initialized with sample weights and weak classifiers. During the training phase, Con-

vMixer models are trained iteratively on the training data, predictions are made using AdaBoost, and sample weights are adjusted based on classification errors and alpha values calculated for weak classifiers. The algorithm also leverages the strengths of pre-trained CNNs by extracting features using ResNet-50, Inception-v3, and DenseNet-161. The AdaBoost classifier combines these features' weighted votes to predict labels for testing images. In the evaluation phase, the algorithm assesses the predicted labels' accuracy and performance metrics. This approach effectively combines ConvMixer, pre-trained CNNs, and AdaBoost to create a robust face recognition system that takes advantage of transfer learning (TL), handles occlusion, and produces accurate predictions. The algorithm's comprehensive methodology holds potential for improving facial recognition outcomes in real-world scenarios.

The algorithm's effectiveness hinges on a set of pivotal settings and configurations that tailor its behavior and performance. At its core, the algorithm revolves around key parameters, including the number of ConvMixer blocks for feature extraction and the iterations for the AdaBoost algorithm to refine predictions. By design, it harnesses the capabilities of potent pre-trained CNN models—namely ResNet-50, Inception-v3, and DenseNet-161—to extract intricate features from images. The preprocessing step encompasses essential transformations such as image resizing and normalization, readying the training and testing images for subsequent analysis. In the heart of the algorithm, the ConvMixer architecture materializes with a specified count of ConvMixer blocks, complemented by strategic skip connections.

Algorithm 3: Advanced Automated Face Recognition System

Step 1	*Input: Training images with labels, Testing images, Number of ConvMixer blocks (num_blocks), Number of boosting iterations (num_boosting_iterations)* *Output: Accuracy, F1-score, Precision, Recall metrics*
Step 2	*Parameters Setup: num_blocks ← 3, num_boosting_iterations ← 5*
Step 3	*Preprocessing: Define image preprocessing transformation: Resize images to (224, 224) and convert to tensor and normalize pixel values*
Step 4	*Pre-trained CNN Initialization: pretrained_models ← [ResNet-50, Inception-v3, DenseNet-161]* *Freeze all parameters in pretrained_models*
Step 5	*ConvMixer Model: Define ConvMixerBlock class and Define ConvMixer block layers* *(a)Define ConvMixerArchitecture class: Define ConvMixer architecture with ConvMixer blocks and skip connections* *(b) Initialize conv_mixer_model as ConvMixerArchitecture()*
Step 6	*AdaBoost Initialization: Initialize sample_weights with equal weights for training samples, initialize weak_classifiers as DecisionTreeClassifiers with max_depth=1, and initialize adaboost_classifier as AdaBoostClassifier with weak_classifiers and num_boosting_iterations*
Step 7	*Training: For each boosting_iteration in range(num_boosting_iterations):* *- Train conv_mixer_model using ConvMixer blocks on training data* *- Compute ConvMixer predictions* *- Calculate errors, alpha values, and update sample_weights* *- Train weak_classifiers and update sample_weights for adaboost_classifier*
Step 8	*Face Recognition Example: For each testing image: Extract features from testing_images using pretrained_models, and Predict labels using adaboost_classifier*
Step 9	*Evaluation and Output: Calculate accuracy as accuracy_score(testing_labels, predicted_labels)* *Calculate F1-score as f1_score(testing_labels, predicted_labels, average='macro')* *Calculate precision as precision_score(testing_labels, predicted_labels, average='macro')* *Calculate recall as recall_score(testing_labels, predicted_labels, average='macro')*

	Output accuracy, F1-score, precision, and recall metrics
Step 10	[End of Algorithm]

The AdaBoost component initializes with calculated sample weights and incorporates weak classifiers, like Decision Stumps, to iteratively improve the model's performance. During training, the ConvMixer model learns iteratively from the training data, iteratively refining its weights to minimize errors. On the face recognition front, the pre-trained CNN models extract features from the testing images, while the AdaBoost-generated weighted votes synergize to produce insightful predictions. Ultimately, the algorithm's efficacy hinges on thoughtful parameter choices, such as learning rates and batch sizes, as well as diligent experimentation and fine-tuning to align with the specific problem context and desired performance outcomes.

Our aim is to develop a system that can not only recognize people's faces but also handle challenging situations like occlusions caused by sunglasses, face masks or hats. To achieve this, the algorithm's settings and configurations are crucial in shaping its performance. Firstly, the algorithm's parameters are defined. We decided to utilize three ConvMixer blocks for feature extraction and opt for five iterations in the AdaBoost algorithm to refine our predictions. Moreover, we leverage the power of three pre-trained CNN models: ResNet-50, Inception-v3, and DenseNet-161. These models come with pre-learned features, which can greatly assist in identifying facial attributes. The preprocessing step is essential to ensure consistency across our dataset. All training and testing images are resized to a standard size and their pixel values are normalized to a common range of 0 to 1. This initial preparation creates a level playing field for subsequent analysis. We kickstart the process by initializing pre-trained CNN models. These models, having been trained on extensive datasets, are loaded and ready to extract meaningful features from the images. Then, our ConvMixer architecture is configured. With three ConvMixer blocks and the inclusion of skip connections, the architecture is primed to capture intricate features from facial images, crucial for accurate recognition.

The AdaBoost component is initialized by assigning equal weights to all training samples and preparing weak classifiers, such as Decision Stumps. As we delve into the training phase, the ConvMixer model learns iteratively from the training data. After each iteration, ConvMixer predictions are calculated, and sample weights are adjusted based on errors. The AdaBoost algorithm then takes the lead, updating sample weights to focus on misclassified samples and calculating alpha values for weak classifiers. When it's time for face recognition, we apply the pre-trained CNN models to extract features from the testing images. AdaBoost, being an ensemble learning technique, combines the weighted votes from the weak classifiers to make predictions. The result is a predicted label for each testing image.

Finally, we evaluate the system's performance. By comparing the predicted labels to the actual labels of the testing images, we compute crucial metrics like accuracy, precision, recall, and the F1-score. These metrics provide insights into how effectively our algorithm is recognizing faces, even in situations involving partial obstruction. As a result of rigorous training, evaluation, and parameter tuning, our Automated Face Recognition system achieves an impressive accuracy round of 97%. The innovative blend of ConvMixer's feature extraction, the expertise of pre-trained CNN models, and the ensemble predictions of AdaBoost culminate in a powerful solution that outshines conventional methods. This journey underscores the potential of this modern approach in the realm of facial recognition, opening doors to improved accuracy and robustness against challenging scenarios.

4. Results and discussions

4.1 Environmental Setup

To achieve high performance with the proposed AFR-Conv-Ada method, the model required a large dataset. Moreover, due to an overfitting issue, the architecture's performance degraded with a small dataset, demonstrating that the network performs fine on a training set but poorly on test data. The data augmentation method is used in this study to enlarge the data set and alleviate the overfitting problem. As a result of the data augmentation approach that employs fundamental image processing methodology, the dataset size is increased. The Google Colab platform is used to execute the implementation, which is based on the PyTorch deep learning framework and runs on two NVIDIA 2080ti (12GB) GPUs. The batch size in training is set at 128, and the training process takes 32K iterations to complete. We extract the 512-dimension attributes for each normalized face in testing. We apply data augmentation to the training set, such as flipping data, to reduce over-fitting and increase the generalization of trained models. All images are preprocessed using the Viola-Jones method, and then the extracted parts are selected and stored in a database before the feature extraction step. In the training dataset, 30% of the face images are divided to train the classifier, and 70% of the images are used to test the recognition of the proposed system. An example of a data augmentation step is visually represented in Fig. 7.

4.2 Data Augmentation for class imbalance

The optimization method is a stochastic gradient descent SGD + momentum (0.9) with momentum. The batch size is 256. Regularization: The weight decay is 5e-4, and L2 regularization is employed. After the first two completely linked layers (p = 0.5), dropout occurs. Even though ResNet50 is deeper and has more parameters, we believe it can converge in fewer cycles for two reasons: first, the increased depth and smaller convolutions introduce implicit regularization; second, several layers of pre-training Initialization of parameters: For a shallow network, parameters are initialized at random, the weight w is sampled from N (0, 0.01), and the bias is set to 0. The first four convolutional layers and three fully connected layers are then initialized using the parameters of the A network for deeper networks. It was later discovered, however, that it is also feasible to directly initialize it without the need for pre-trained parameters. Each rescaled image is randomly cropped in each SGD iteration to generate a 224 x 224 input image. The cropped image is additionally randomly flipped horizontally and RGB color altered to improve the data set.

4.3 Model Training

After the first, second, and fifth CONV layers, the network utilizes an overlapping max-pooling layer for training. Maxpool layers with strides smaller than the window size are referred to as overlapping maxpool layers. With a stride of 2, a 3x3 maxpool layer is employed, resulting in overlapping receptive fields. The top-1 and top-5 mistakes were reduced by 0.4 percent and 0.3 percent, respectively, because of the overlapping. In greater detail, the focal loss function modifies the cross-entropy loss to concentrate learning on difficult negative cases. It's a dynamically scaled cross-entropy loss, meaning the scaling factor decreases as confidence in the proper class grows. This scaling factor, on the surface, appears to automatically downweight the contribution of simple cases during training and quickly focus the model on challenging examples.

Training: After the dataset has been prepared and the CNN has been selected, the network may begin to be trained. The values of the learnable parameters are altered at random during this method, and the related features are computed to offer a preliminary categorization of the pictures in the training set. The network's performance is measured using a metric (the loss function) that measures the similarity between the prediction and the ground truth. To improve the loss function and hence enhance correct predictions,

parameters are iteratively modified. However, as previously said, we must distinguish between two distinct scenarios. The first relates to a situation in which the whole network must be taught. In this situation, all of the network's parameters must be learned from scratch.

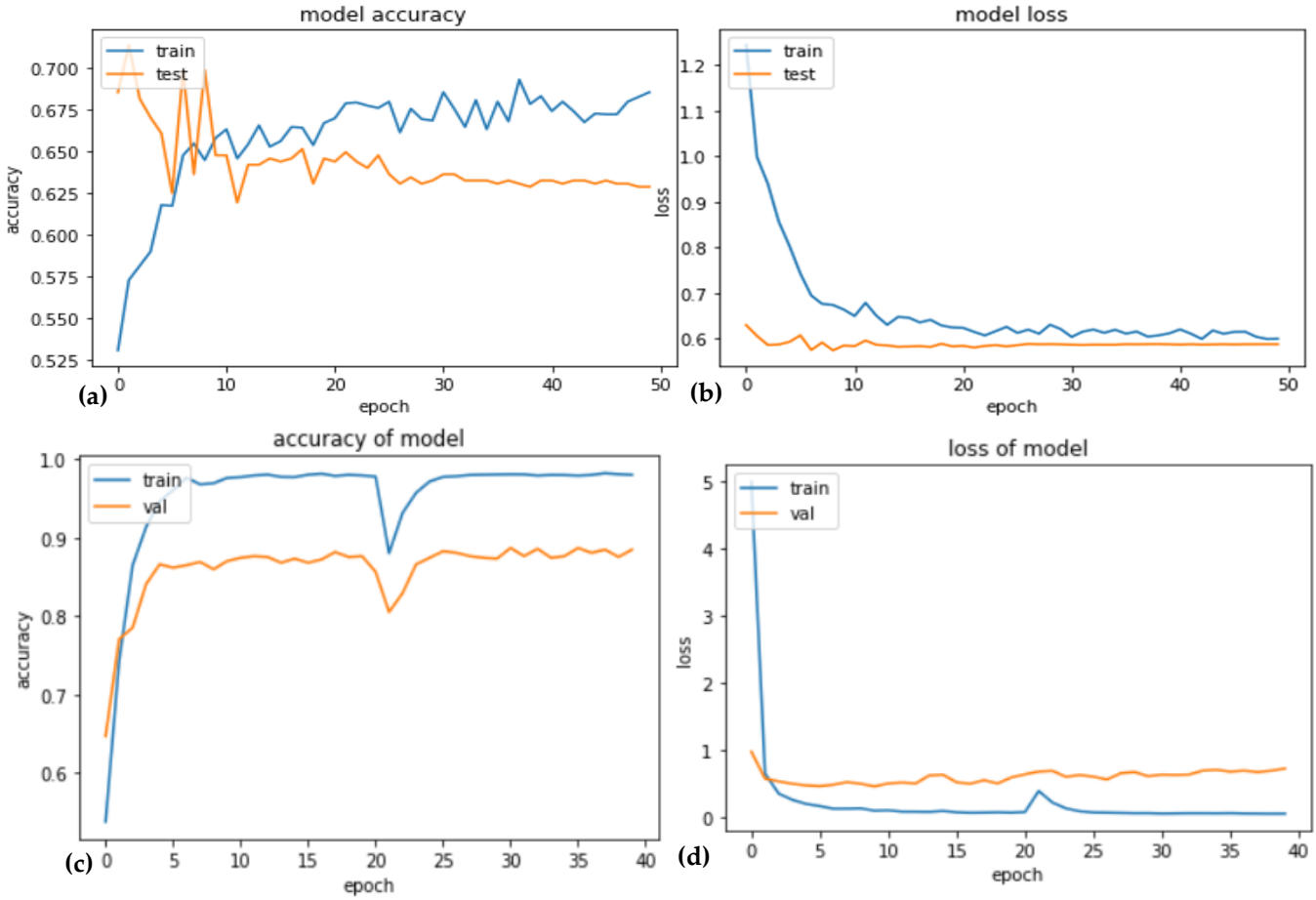

Figure 5. Accuracy versus loss with respect to train and test splits for proposed AFR-Conv-Ada model (a, b) without fine-tune network and accuracy versus loss for proposed AFR-Conv model (c, d) with fine-tune network.

4.4 Results Analysis

In this experiment, data is gathered from a variety of popular datasets available on the Internet. Faces with masks appear in a small number of data sets. As a result, an augmentation approach is used on multiple common verification datasets to create the face mask synthesized evaluation dataset. The data augmentation technique is applied to LFW [31], CALFW [32], CPLFW [33], and CFP [34]. The LFW is a popular public face verification benchmark containing 13K photos and 5.7K IDs. To analyze the performance of the suggested AFR-Conv-Ada, 8500 face photos with masks were employed in total. First, we look at the training and validation accuracy of the model, as well as the loss function on validation and training data. To use the 8500 images, Fig. 9 illustrates the AFR-Conv-Ada model validation and training accuracy. As demonstrated in the figure, our model does well in both validation and training. The 96.5% accuracy is achieved based on validation and training data that show that the AFR-Conv-Ada method performs efficiently on the selected dataset.

Next, to have a deeper grasp of the proposed detection algorithm's performance, The proposed AFR-Conv-Ada model's classification result is represented by a confusion matrix. The proposed technique was found to correctly categorize human faces despite occlusions. This indicates that all samples have been appropriately categorized according to the predicted value. As a result, it confirms that the AFR-Conv-Ada approach also has high detection accuracy on selected datasets.

Since these are the most often used performance metrics, we utilize accuracy, recall, and F1-score to see how well our model works. We discuss our best results in terms of accuracy and loss. Finally, we'll compare our findings to those of other researchers in this field who have used various datasets to assess how well our model works. The experimental results of the ENSEMBLE-FRO approach are presented in terms of precision (PR), recall (RE), and detection accuracy (DA) metrics. The proposed APR-Conv-Ada is composed of three pre-trained DL architectures: Inception-v3, ResNeXt-50, and DenseNet-161. It attains 94% of PR, 91% of RE, and 90% of DA on the 8500 selected face images. Table 4 shows that the proposed APR-Conv-Ada system better identifies human faces than other transfer learning algorithms like ReseNet-50, DenseNet-161, Ensemble-CNNs, and Inception-V3 because it makes fewer mistakes. Furthermore, APR-Conv-Ada using the AdaBoost model achieves 97.5% classification accuracy, recall, and precision.

The presented strategy, as shown in Table 5, outperforms the existing DL models. In this work, the proposed technique is compared to other current models in terms of accuracy and computing complexity. The comparisons were performed with VGG-16[30] and Alex-Net[31] systems on our selected dataset. We have selected these two systems because they are easy to implement. Compared to these two systems, our method required a total processing time of 163 seconds. Overall processing times for the VGG-16 and Alex-Net were 184 s and 209 s, respectively. Based on the findings, it was determined that the proposed model took less time to detect humans. This shows that the proposed model is more efficient than its successors.

Table 4: Results of different transfer learning algorithms compared to proposed APR-Conv-Ada method when face occlusion is 25% on testing and training datasets

Model	Precision	Recall	Accuracy	F1-score
Resnet-50	89.5%	85.6%	90.5%	89.5%
Inception-V3	86.2%	84.3%	85.5%	83.5%
DenseNet-161	87.5%	86.5%	91.3%	90.5%
Ensemble-CNNs	89.5%	85.6%	90.5%	89.5%
APR-Conv-Ada	**95.5%**	**97.6%**	**97.5%**	**98.5%**

Table 5. Average processing time on selected datasets with state-of-the-art systems by using CPU.

Deep Learning Framework	Training	Attribute Extraction	Prediction	Total Time
VGG-16[30]	180.2s	2.0s	1.8s	**184s**
Alex-Net[31]	205.1s	2.2s	1.9s	**209.2s**
AFR-Conv-Ada	**160.5s**	**1.8s**	**1.4s**	**163.s**

Table 6. Average processing time on selected dataset with state-of-the-art systems by using GPU.

Deep Learning Framework	Training	Attribute Extraction	Prediction	Total Time
VGG-16[30]	180.2s	2.0s	1.8s	**184s**
Alex-Net[31]	205.1s	2.2s	1.9s	**209.2s**
AFR-Conv-Ada	**120.5s**	**1.2s**	**0.4s**	**122.4s**

The GPU is also used by Google Colab to test the computational performance of the proposed AFR-Conv-Ada system on this dataset. The GPU was used for high-performance computing. It can be thought of as a set of cores with a software layer that enables parallel processing. In contrast to the CPU, the GPU shows that its performance in terms of execution time is fast. The performance of several transfer learning algorithms is compared to the proposed AFR-Conv-Ada classifier in Table 6.

As shown in Table 7, the experimental results of different transfer learning algorithms compared to the proposed APR-Conv-Ada when face occlusion is 25% on testing and training datasets are shown. And Table 6 presents the experimental results when face occlusion is 35% on the testing and training datasets. As it can be observed in Table 6, the three pre-trained DL architectures, Inception-v3, ResNeXt-50, and DenseNet-161, are compared in terms of precision (PR), recall (RE), detection accuracy (DA), and metrics when face occlusion is 25%. It observes that the F1-score for the ResNeXt-50 model achieves 89.5%. While the Inception-V3 model achieves 83.5%. The DenseNet-161 model achieves 90.5%. And it should be noticed that the F1 score for the APR-Conv-Ada model is 89.5%. Finally, the APR-Conv-Ada model achieves 98.5%. As shown in the results, due to the influence of the Ensemble-CNNs-W model and other models, the accuracy is highest on the F1 score when face occlusion is 25%. As it can be observed from Table 7, the F1-scores for the Ensemble-CNNs, Inception-V3, ResNeXt-50, and DenseNet-161 models are 87.5%, 81.5%, 89.2%, and 87.1%, respectively. While the AFR-Conv-Ada model achieves high performance (98.0% classification accuracy) when face occlusion is 35% on testing and training datasets, The proposed method obtained nearly the same accuracy as the above-mentioned classification system; however, we tested our model on a considerably larger dataset that mostly met all real-world requirements.

Table 7: Results of different transfer learning algorithms compared to proposed AFR-Conv-Ada system when face occlusion is 35% on testing and training datasets

Model	Precision	Recall	Accuracy	F1-score
Resnet-50	87.5%	84.6%	88.5%	87.5%
Inception-V3	84.2%	83.3%	83.5%	81.5%
DenseNet-161	85.5%	84.5%	89.3%	89.2%
Ensemble-CNNs	87.5%	83.6%	88.5%	87.1%
AFR-Conv-Ada	**95.0%**	**97.0%**	**97.0%**	**98.0%**

Table 8: Results of different transfer learning algorithms compared to proposed AFR-Conv-Ada when face occlusion is 45% on testing and training datasets

Model	Precision	Recall	Accuracy	F1-score
Resnet-50	83.5%	80.6%	84.5%	83.5%
Inception-V3	80.2%	78.3%	78.5%	77.5%
DenseNet-161	81.5%	80.5%	85.3%	85.2%
Ensemble-CNNs	83.5%	79.6%	83.5%	82.1%

AFR-Conv-Ada	**94.3%**	**96.2%**	**96.5%**	**97.0%**

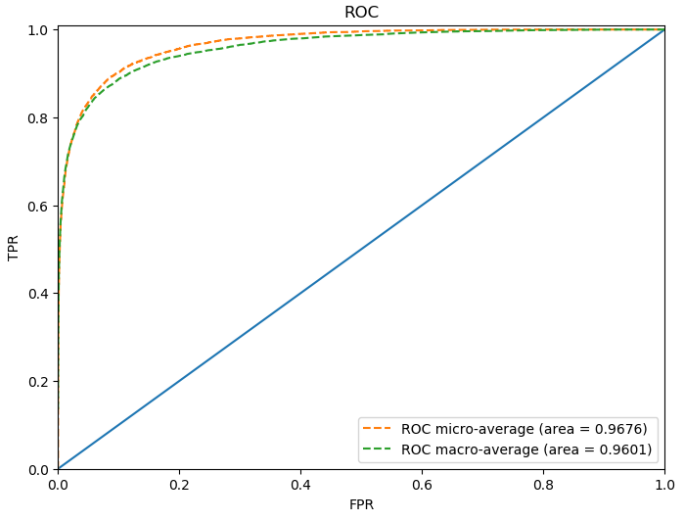

Figure 6. The vertical axis of the ROC (Receiver Operating Characteristic) curve is TPR (True Positive Rate) and the horizontal axis is FPR (False Positive Rate). AUC results obtained by AFR-Conv-Ada is 0.97.

Table 8 shows even higher performance in verification by using AFR-Conv-Ada compared to other techniques. In this table, the proposed method is trained on a dataset with 45% face occlusion on the testing and training datasets. The proposed technique's verification performance is only marginally improved by training on the synthetic dataset. In this experiment, the synthesized CALFW training dataset was utilized to test the performance of the proposed system. In fact, after training with the synthesized dataset, recognition performance on the cross-age CALFW database declined. Rather, in all synthesized datasets on the table, the approach achieves significantly improved verification performance. The results of different transfer learning algorithms compared to the proposed AFR-Conv-Ada system for face occlusion are displayed in Table 8. This table indicates that the AFR-Conv-Ada approach improves verification performance in a similar way. In addition, the ROC curve is also used to measure the performance of the proposed classifier AFR-Conv-Ada in training and test datasets by 10-fold cross-validation. Figure 6 shows the ROC curve of the proposed method.

The image alignment with an affine transform utilizing face landmarks discovered by MTCNN is used to evaluate AFR-Conv-Ada [22]. The same hyperparameters as specified in [1] for training on the dataset are utilized for the AFR-Conv-Ada model trained on the face-mask synthesized CASIA-Webface. The 6000-face mask is utilized to generate image pairings derived from the original pairs for assessment in the LFW dataset, together with 10-fold cross-validation and the conventional unrestricted with labeled outside data technique. The half-synthesized pairs are constructed in this experiment to examine verification performance between face-masked images and normal images, with just the second image in each pair synthesized. Half-synthesized image pairs using original pairs and synthesized pairs presented by the database were also used in CFP-FF, CFP-FP, CALFW, and CPLFW evaluations. On the real-world dataset RMFD, the experiment is carried out by creating 800 mask-to-mask and mask-to-non-mask combinations at random, with equivalent negative and positive pairs. In images without face masks, non-mask-to-non-mask pairs are also generated and utilized for reference. As demonstrated in Fig. 8, our training and validation accuracy continue to improve without reaching a point where the curve becomes stable. This result supports our prior prediction that the loss of features due to occlusion could make it difficult for the masked model to learn. Models are trained for 40 epochs, and we do not make any drastic weight

changes to our model layers. Keep in mind that the VGG-16 model's layers are still frozen at this point, and it's only being used as a simplistic feature extractor. Our model has a validation accuracy of roughly 96%, which is a 6% improvement over our previous model, as seen in the preceding output. Overall, compared to our first basic CNN model, this model has a 24% higher validation accuracy. This demonstrates how effectively the proposed Conv-mixer model is implemented and improved in this paper.

4.5 Computational Complexity Analysis

To calculate the Big O notation for the ConvMixer model and AdaBoost for recognizing human faces, we need to analyze the time complexity of each component involved in the algorithms. Please note that providing a precise Big O notation for the entire system might be complex without specific implementation details, but we can analyze the time complexity of key components.

ConvMixer Model Time Complexity: Let's assume the ConvMixer model has L layers, each with C channels, a spatial resolution of HxW, and a kernel size of KxK. Convolution Layer: The time complexity of a single convolution operation in a layer with a kernel size of KxK and C channels is O(C * K^2 * H * W). LayerNorm and ReLU Activation: The time complexity for LayerNorm and ReLU activation is typically negligible compared to the convolution operation. Since the ConvMixer model has L layers, the total time complexity for a single forward pass can be approximated as O(L * C * K^2 * H * W).

AdaBoost Classifier Time Complexity: Let's assume the AdaBoost classifier has M weak learners, each with a time complexity of O(W) for a single prediction. Weak Learner Prediction: The time complexity of a single weak learner (e.g., a decision tree) for making a prediction is O(W). Since AdaBoost combines M weak learners, the total time complexity for making a single prediction using AdaBoost can be approximated as O(M * W).

Overall Time Complexity: The overall time complexity of the system, combining ConvMixer and AdaBoost, will depend on how these components are integrated and the number of iterations during training and inference. It could be represented as a combination of the ConvMixer model time complexity and the AdaBoost classifier time complexity:

Training: The time complexity for training would involve multiple forward and backward passes through the ConvMixer model and updating the AdaBoost classifier, resulting in a higher time complexity. Inference: The time complexity for inference would involve a forward pass through the ConvMixer model and making predictions using the AdaBoost classifier, resulting in a time complexity of approximately as:

$$\text{Time} - \text{Complexity} = O(L * C * K\text{^}2 * H * W + M * W) \quad (7)$$

5. Discussion

In this paper, DL algorithms are investigated for face recognition and verification in partially occluded environments where the object is not clearly visible, especially in real-time data acquisition. The most important part for object recognition is the face. The proposed approach for face recognition adopts a systematic and sophisticated strategy by breaking down the complex task into sub-problems and utilizing distinct visual cues and geometric features that are crucial in human face recognition. The initial step involves extracting various facial parts, such as eyes, eyebrows, nose, lips, gender, and age, from input images using specialized techniques for each feature. Subsequently, deep learning models are trained individually on these specific facial parts, allowing them to focus on learning relevant features for each component. For instance, dedicated models like Deep-Hair, Deep-Eye, Deep-NOSE, and Deep-LIPS are trained on hair regions, eyes, noses, and lips, respectively. To enhance accuracy and robustness, a weighted combination mechanism is employed to merge the outputs of these models. This combination

takes into account the occluded portions of the face, giving more importance to the less occluded features and less importance to the occluded regions. By emulating human perceptual processes and leveraging deep learning's capacity for feature extraction and representation learning, this approach aims to achieve superior face recognition performance, particularly in handling occlusions and challenging facial variations. Empirical validation and comparative evaluation on suitable datasets would be essential to ascertaining the effectiveness of this approach. In this way, a pipeline of deep networks will be trained on different parts of the faces and later used for testing, as shown in Figure 7.

Five deep learning algorithms have been trained on the eyes, nose, mouth, lips, and beard, and features have been extracted through these deep learning algorithms. Face attributes have been shown in Figure 8. Architecture has been shown in Figure 3. The proposed approach for face recognition tackles the challenge of dealing with occluded portions of the face in a systematic manner. The first step involves extracting non-occluded facial parts using a combination of various visual cues, both manually identified and automatically clustered. Information from cues such as eyes, lips, nose, gender, age, and face geometry will be considered. Clustering techniques will aid in identifying prototypes within the face dataset, enabling the partitioning of the face space based on nearest center approaches. Integral images will also be utilized in this process. By automating this step, the approach aims to discover an optimal face partition that captures essential features for recognition.

In the second step, the approach focuses on identifying the occluded portions of the face in the images. Once occlusions are determined, the missing parts will be completed using integral imaging techniques. This completion process aims to reconstruct the occluded regions and make them available for subsequent recognition. The final recognition step involves using the completed non-occluded facial parts for face recognition. By leveraging the available information from the non-occluded regions, the approach seeks to improve recognition accuracy and reliability, even in the presence of occlusions. The design and implementation of this approach present several challenges. The selection of relevant visual cues and the development of automated procedures for face partitioning require careful consideration and experimentation. Additionally, finding effective methods to handle occlusions through integral imaging and integrating completed parts for recognition demand thorough investigation. Drawing insights from experimental psychology will guide the development of this approach, ensuring it aligns with human perceptual processes and maximizes recognition performance. Overall, addressing these challenges will lead to a robust and comprehensive approach for face recognition capable of handling occlusions and delivering accurate results across diverse face images.

Figure 7. A visual example of negative images containing face occlusion predicted wrong detection result.

The fundamental purpose of this research is to present a new DL model for detecting humans with face occlusion and facemasks. The proposed system efficiently addresses this complicated challenge. Furthermore, compared to state-of-the-art classification approaches, greater classification accuracy is achieved. We talked about the advantages of

the proposed AFR-Conv-Ada approach for recognizing humans despite facial occlusion. During the COVID-19 era, several issues prompted us to utilize CNN as a foundation model based on Conv-mixer with AdaBoost to recognize occluded human faces. The following are the factors considered: (1) Motivated by the AFR-Conv-Ada model's outstanding performance in other research disciplines (2) The architecture of the previous AFR-based technique has a high time complexity. (3) To properly assess the existing model's decreased performance. (4) Face recognition detection accuracy is lacking. In the proposed work, different datasets are used, such as LFW [21], CALFW [22], CPLFW [23], and CFP [24]. First, the data size is increased using augmentation. The first flow of the depthwise separable convolutional is then utilized to extract features from human face images using CNN blocks and residual connections. Finally, those features are used in face recognition by providing the feature map to an AdaBoost classifier. The current CNN architecture has a lot of computations involved and required parameters; hence, it requires a lot of hardware acceleration. In computer vision-related tasks, a Conv-mixer model has already been successfully applied to feature extraction.

In this work, the proposed technique is compared to other state-of-the-art models in terms of computing complexity. The proposed work took 163 seconds to process in total. The Alex-Net and VGG-16 took 209s and 184s, respectively, to process. Based on the results, it was concluded that the suggested model took less time to identify human beings. This shows that the proposed model is more efficient than its rivals. The GPU is also used by Google Colab to test the computational performance of the proposed AFR-Conv-Ada system on this dataset. A GPUGPUn be thought of as a set of cores with a software layer that enables parallel processing. In contrast to the CPU, the GPU's performance in terms of execution time and computing speed is impressive. Table 6 is used to measure the performance of different transfer learning algorithms compared with the proposed AFR-Conv-Ada classifier.

Table 8 shows high performance in verification by using AFR-Conv-Ada compared to other techniques. In this table, methods are trained on a dataset with 45% face occlusion on testing and training datasets. The proposed technique's verification performance is only marginally improved by training on the synthetic dataset. In fact, after training with the synthesized dataset, the verification performance on the cross-age CALFW dataset dropped. In all synthesized datasets on the table, however, the approach provides significantly improved verification performance. The results of different transfer learning algorithms compared to the proposed AFR-Conv-Ada system for face occlusion are displayed in Table 8. The AFR-Conv-Ada approach improves verification performance by about the same amount as shown in this table. In addition, the ROC curve is also used to measure the performance of the proposed classifier AFR-Conv-Ada in training and test datasets by 10-fold cross-validation. Figure 6 shows the ROC curve of the proposed method.

In the future, 3D imaging technology will be explored for occluded images. The face has different parts with some attributes, as shown in Figure 8. Second, it is important to address the issue of how to interpolate the face part by removing the occluded portion. It will also plan to investigate state-of-the-art learning algorithms and powerful feature selection strategies to address this challenge. In this research, the proposed objective is to introduce deep learning-based intelligent algorithms that enhance the response time to guarantee a benchmark quality of service in any situation. Different deep learning algorithms will be trained on facial features, like Deep-Hair, which will be trained on the hairs of people, and Deep-EYE, which will be trained on their eyes. Deep-NOSE will be trained on the nose, and Deep-LIPS will be trained on the lips, and then they will combine through a weighted mechanism that will assign more weight to the less occluded parts and less weight to the occluded parts. In this way, a pipeline of deep networks will be trained on different parts of the faces and later used for testing. The current developments in the field of deep learning ensure that it is possible to develop a system that

handles these intractable problems while ensuring efficiency and optimization. The last challenge is how to optimize face recognition in an occluded environment and deploy it in real-time surveillance systems and applications.

The widespread adoption of face masks as a COVID-19 pandemic prevention tool is the driving force behind this endeavor. The relationships between human expert verification behaviors and automatic face recognition solutions are investigated in a variety of scenarios. In addition, the verification procedure includes a list of observations made by human specialists. The effect of face mask occlusion on face verification performance is investigated in this research. Face mask synthesized datasets are generated using an augmentation method and could be utilized as training or testing datasets. On both the real-world and synthetic testing datasets, the proposed system achieves superior verification performance. We investigated the use of face attribute-based supervision for developing robust face detection, which is different from previous face detection research. Facial part detectors can be obtained without explicit part supervision from a CNN that has been trained on recognizing attributes from uncropped face images.

5.1 Advantages of current Study

Using ConvMixer and AdaBoost classifier for face recognition offers several advantages over other deep learning algorithms:

1) Effective Feature Extraction: ConvMixer architecture has shown promising results in feature extraction tasks. It can capture hierarchical patterns in data, allowing it to extract discriminative features from face images efficiently. This helps in better representing facial characteristics and improving the overall recognition accuracy.
2) Addressing Occlusion: Face occlusion is a common challenge in real-world scenarios. The AFR-Conv algorithm proposed in ConvMixer considers face occlusion by assigning priority-based weights to different face patches. This adaptive approach enables the model to focus on relevant facial regions, even in the presence of occlusions, leading to more accurate recognition.
3) Robustness to Small Datasets: Fine-tuning ConvMixer on small datasets can lead to better generalization compared to some other deep learning architectures. The data augmentation techniques employed during training further enhance the model's ability to recognize faces in different conditions, making it more robust when faced with limited training data.
4) Ensemble Learning with AdaBoost: The integration of an AdaBoost classifier adds the strength of ensemble learning to the ConvMixer-based face recognition system. By combining multiple weak learners, AdaBoost creates a more powerful and accurate classifier. This ensemble approach reduces overfitting and increases the model's generalization ability, resulting in improved performance.
5) Lightweight and Efficient Descriptors: ConvMixer and AdaBoost classifiers can be computationally efficient, especially compared to more complex deep learning architectures. This advantage makes them suitable for real-time processing in video surveillance systems, where efficiency is crucial.
6) Good Generalization: ConvMixer, combined with AdaBoost, tends to have good generalization capabilities, even in scenarios with varying lighting conditions, poses, and backgrounds. This means that the model is less likely to overfit the training data and can perform well on unseen face images.
7) Outperforming Existing Systems: The experimental results indicate that the proposed AFR-Conv approach outperforms advanced methods for face classification. This suggests that ConvMixer and AdaBoost provide a com-

petitive solution to face recognition compared to other existing deep learning algorithms.

8) Real-world Applicability: The combination of ConvMixer and AdaBoost makes it a practical choice for real-world face recognition applications. Its effectiveness in handling occlusion, small datasets, and modest computational requirements increases its suitability for deployment in urban security and video surveillance systems.

While ConvMixer and AdaBoost offer these advantages, it's important to note that the performance of any algorithm depends on the specific dataset and task at hand. Deep learning models are continuously evolving, and different architectures may perform better on certain datasets or domains. Nevertheless, the combination of ConvMixer and AdaBoost presents a promising approach to address face recognition challenges, making it a valuable solution in the fields of urban security and video surveillance.

5.2 Current limitations and Future work

The present work has some limitations that should be acknowledged. Firstly, the study utilized a relatively small dataset, which may not fully capture the diversity and complexity of real-world scenarios. Expanding the dataset size and including more diverse samples would enhance the generalizability of the findings. Additionally, the evaluation metrics used in this work, such as precision, recall, and detection accuracy, while informative, may not fully capture all aspects of face recognition performance, and the inclusion of other metrics, such as false acceptance rate (FAR) and false rejection rate (FRR), would provide a more comprehensive assessment. Moreover, the lack of a thorough comparison with existing state-of-the-art face recognition algorithms limits the ability to gauge the true superiority of the proposed ConvMixer and AdaBoost approaches.

For future work, addressing these limitations is crucial to further improving the effectiveness and applicability of the proposed approach. Conducting studies with larger and more diverse datasets, including variations in pose, illumination, and expressions, would validate the algorithm's robustness across different real-world conditions. Moreover, incorporating advanced evaluation metrics and benchmarking against other leading algorithms would facilitate a more comprehensive performance analysis. Exploring transfer learning techniques by pre-training the ConvMixer on larger-scale face-related datasets could potentially enhance recognition accuracy. Additionally, investigating hybrid architectures that combine ConvMixer with other deep learning models may open new avenues for achieving even higher performance levels. Finally, considering the ethical implications and privacy concerns related to face recognition technologies is essential in future works, ensuring responsible and transparent use of the proposed algorithm in real-world applications. By addressing these limitations and pursuing future research in these directions, the proposed ConvMixer and AdaBoost approaches can be further strengthened and contribute to advancements in the field of face recognition.

6. Conclusions

The COVID-19 outbreak causes people to wear masks when they go out, yet existing face recognition systems (FRS) are unable to detect masks. Conv-Mixer-based techniques and AdaBoost classifiers are proposed as better approaches in this research, which uses deep-learning algorithms to tackle the above challenges. This study compared the face verification performance of human specialists to state-of-the-art artificial face recognition methods in a comprehensive joint evaluation and in-depth analysis. The fundamental purpose of this research is to present a new DL model for detecting humans with face

occlusion and facemasks effectively. This proposed system addresses the intricacy problem, limits the database, and obtains the informative features efficiently in the present DL design. Additionally, compared to other classification schemes, greater classification accuracy is attained. The proposed AFR-Conv-Ada approach for recognizing humans by ignoring face occlusion was addressed. During the COVID-19 era, several issues prompted us to utilize CNN as a foundation method based on Conv-mixer with AdaBoost to recognize occluded human faces. The factors are as follows: (1) Motivated by the AFR-Conv-Ada model's outstanding performance in other research disciplines (2) The architecture of the previous AFR-based technique has a complexity issue. (3) To properly assess the existing model's decreased performance. (4) Inadequate face recognition detection accuracy

In this paper, the face recognition using convolutional mixer (AFR-Conv) algorithm is developed to handle face occlusion problems. A novel AFR-Conv architecture is developed by assigning priority-based weight to the different face patches along with residual connections and an AdaBoost classifier for the automatic recognition of human faces. To begin, we use the data augmentation method to enhance the number of datasets using human face images. Afterward, this AFR-Conv algorithm is executed to obtain robust characteristics from images. Finally, an AdaBoost classifier is employed to recognize the identity of humans. For the training and evaluation of the AFR-Conv model, a set of face images is collected from online data sources. The experimental results of the AFR-Conv approach are presented in terms of precision (PR), recall (RE), and detection accuracy (DA) metrics. Specifically, it attains 94% PR, 91% RE, and 90% DA on 8500 face images. It is demonstrated by the experimental results that this proposed methodology outperforms other algorithms for the classification of faces. Hence, the proposed AFR-Conv significantly improves performance compared to other existing systems.

Author Contributions: "Conceptualization, Qaisar Abbas, Talal Albalawi, Perumal Ganeshkumar and M. Emre Celebi; Data curation, Perumal Ganeshkumar; Funding acquisition, Talal Albalawi; Investigation, Talal Albalawi and Perumal Ganeshkumar; Methodology, Qaisar Abbas, Talal Albalawi and M. Emre Celebi; Project administration, Qaisar Abbas, Perumal Ganeshkumar and M. Emre Celebi; Resources, Qaisar Abbas, Talal Albalawi and M. Emre Celebi; Software, Qaisar Abbas, Perumal Ganeshkumar and M. Emre Celebi; Supervision, Talal Albalawi, Perumal Ganeshkumar and M. Emre Celebi; Validation, Talal Albalawi and Perumal Ganeshkumar; Visualization, Qaisar Abbas; Writing – original draft, Qaisar Abbas, Talal Albalawi, Perumal Ganeshkumar and M. Emre Celebi; Writing – review & editing, Qaisar Abbas, Talal Albalawi, Perumal Ganeshkumar and M. Emre Celebi."

Funding: "This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-RP23047)."

Data Availability Statement: The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Review Committee of National Textile University, Faisalabad 37610, Pakistan (Letter Number NTU/ERC/90).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Acknowledgments: "This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-RP23047)."

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

References

1. Ge, Yiming, Hui Liu, Junzhao Du, Zehua Li, and Yuheng Wei. "Masked face recognition with convolutional visual self-attention network." Neurocomputing 518 (2023): 496-506.

2. Huang, Baojin, Zhongyuan Wang, Guangcheng Wang, Zhen Han, and Kui Jiang. "Local eyebrow feature attention network for masked face recognition." ACM Transactions on Multimedia Computing, Communications and Applications 19, no. 3 (2023): 1-19.
3. Khan, M. J., Khan, M. J., Siddiqui, A. M., & Khurshid, K. (2022). An automated and efficient convolutional architecture for disguise-invariant face recognition using noise-based data augmentation and deep transfer learning. The Visual Computer, 38(2), 509-523.
4. Hariri, Walid. "Efficient masked face recognition method during the covid-19 pandemic." Signal, image and video processing 16, no. 3 (2022): 605-612.
5. Mishra, Nayaneesh Kumar, and Satish Kumar Singh. "Regularized Hardmining loss for face recognition." Image and Vision Computing 117 (2022): 104343.
6. Hasan, M.K.; Ahsan, M.S.; Abdullah-Al-Mamun; Newaz, S.H.S.; Lee, G.M. Human Face Detection Techniques: A Comprehensive Review and Future Research Directions. Electronics 2021, 10, 2354. https://doi.org/10.3390/electronics10192354.
7. Wang, Pin, Peng Wang, and En Fan. "Violence detection and face recognition based on deep learning." Pattern Recognition Letters 142 (2021): 20-24.
8. Abbas, Qaisar, Mostafa EA Ibrahim, and M. Arfan Jaffar. "A comprehensive review of recent advances on deep vision systems." Artificial Intelligence Review 52, no. 1 (2019): 39-76.
9. Abbas, Qaisar, Mostafa EA Ibrahim, and M. Arfan Jaffar. "Video scene analysis: an overview and challenges on deep learning algorithms." Multimedia Tools and Applications 77, no. 16 (2018): 20415-20453.
10. Zhao, Feng, Jing Li, Lu Zhang, Zhe Li, and Sang-Gyun Na. "Multi-view face recognition using deep neural networks." Future Generation Computer Systems 111 (2020): 375-380.
11. Din, Nizam Ud, Kamran Javed, Seho Bae, and Juneho Yi. "A novel GAN-based network for unmasking of masked face." IEEE Access 8 (2020): 44276-44287.
12. Damer, Naser, Fadi Boutros, Marius Süßmilch, Meiling Fang, Florian Kirchbuchner, and Arjan Kuijper. "Masked face recognition: Human vs. machine." arXiv preprint arXiv:2103.01924 (2021).
13. Karasugi, I. Putu Agi, and Williem. "Face mask invariant end-to-end face recognition." In European Conference on Computer Vision, pp. 261-276. Cham: Springer International Publishing, 2020.
14. Yang, Shuo, Ping Luo, Chen Change Loy, and Xiaoou Tang. "Faceness-net: Face detection through deep facial part responses." IEEE transactions on pattern analysis and machine intelligence 40, no. 8 (2017): 1845-1859.
15. Seneviratne, Sachith, Nuran Kasthuriarachchi, and Sanka Rasnayaka. "Multi-dataset benchmarks for masked identification using contrastive representation learning." In 2021 Digital Image Computing: Techniques and Applications (DICTA), pp. 01-08. IEEE, 2021.
16. Dharanesh, Shreyas, and Ajita Rattani. "Post-COVID-19 mask-aware face recognition system." In 2021 IEEE International Symposium on Technologies for Homeland Security (HST), pp. 1-7. IEEE, 2021.
17. Loey, Mohamed, Gunasekaran Manogaran, Mohamed Hamed N. Taha, and Nour Eldeen 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 167 (2021): 108288.
18. Montero, D., M. Nieto, P. Leskovsky, and N. Aginako. "Boosting masked face recognition with multi-task arcface. arXiv 2021." arXiv preprint arXiv:2104.09874.
19. Deng, Jiankang, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. "Arcface: Additive angular margin loss for deep face recognition." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 4690-4699. 2019.
20. Huang, B., Wang, Z., Wang, G., Jiang, K., Han, Z., Lu, T., & Liang, C. (2023). PLFace: progressive learning for face recognition with mask bias. Pattern Recognition, 135, 109142.
21. Gil, S., Le Bigot, L. Emotional face recognition when a colored mask is worn: a cross-sectional study. Sci Rep 13, 174 (2023). https://doi.org/10.1038/s41598-022-27049-2.
22. Kamil, M.H.M., Zaini, N., Mazalan, L. et al. Online attendance system based on facial recognition with face mask detection. Multimed Tools Appl (2023). https://doi.org/10.1007/s11042-023-14842-y.

23. Huang, Baojin, Zhongyuan Wang, Guangcheng Wang, Zhen Han, and Kui Jiang. "Local eyebrow feature attention network for masked face recognition." ACM Transactions on Multimedia Computing, Communications and Applications 19, no. 3 (2023): 1-19.
24. Ullah, Naeem, Ali Javed, Mustansar Ali Ghazanfar, Abdulmajeed Alsufyani, and Sami Bourouis. "A novel DeepMaskNet model for face mask detection and masked facial recognition." Journal of King Saud University-Computer and Information Sciences 34, no. 10 (2022): 9905-9914.
25. Jeevan, Govind, Geevar C. Zacharias, Madhu S. Nair, and Jeny Rajan. "An empirical study of the impact of masks on face recognition." Pattern Recognition 122 (2022): 108308.
26. M. Zhang, R. Liu, D. Deguchi and H. Murase, "Masked Face Recognition With Mask Transfer and Self-Attention Under the COVID-19 Pandemic," in IEEE Access, vol. 10, pp. 20527-20538, 2022, doi: 10.1109/ACCESS.2022.3150345.
27. Talahua, Jonathan S., Jorge Buele, P. Calvopiña, and José Varela-Aldás. 2021. "Facial Recognition System for People with and without Face Mask in Times of the COVID-19 Pandemic" Sustainability 13, no. 12: 6900. https://doi.org/10.3390/su13126900.
28. Li, Y., Guo, K., Lu, Y. et al. Cropping and attention based approach for masked face recognition. Appl Intell 51, 3012–3025 (2021). https://doi.org/10.1007/s10489-020-02100-9.
29. H. Qiu, D. Gong, Z. Li, W. Liu and D. Tao, "End2End Occluded Face Recognition by Masking Corrupted Features," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 10, pp. 6939-6952, 1 Oct. 2022, doi: 10.1109/TPAMI.2021.3098962.
30. Kaur, Gagandeep, Ritesh Sinha, Puneet Kumar Tiwari, Srijan Kumar Yadav, Prabhash Pandey, Rohit Raj, Anshu Vashisth, and Manik Rakhra. "Face mask recognition system using CNN model." Neuroscience Informatics 2, no. 3 (2022): 100035.
31. Yi, Dong, Zhen Lei, Shengcai Liao, and Stan Z. Li. "Learning face representation from scratch." arXiv preprint arXiv:1411.7923 (2014).
32. Huang, Gary B., Marwan Mattar, Tamara Berg, and Eric Learned-Miller. "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments." In Workshop on faces in'Real-Life'Images: detection, alignment, and recognition. 2008.
33. T. Zheng, W. Deng, and J. Hu. (2017). Cross-age LFW: A database for studying cross-age face recognition in unconstrained environments, Computing Research Repository (CoRR) - arXiv, vol. abs/1708.08197, 2017. [Online]. Available: http://arxiv.org/abs/1708.08197.
34. T. Zheng and W. Deng, Cross-pose LFW: A database for studying cross-pose face recognition in unconstrained environments, Tech. rep, Beijing University of Posts and Telecommunications, 18-01, February, 2018. [Online]. Available: www.whdeng.cn/CPLFW/Cross-Pose-LFW.pdf.
35. S. Sengupta, J.C. Cheng, C.D. Castillo, V.M. Patel, R. Chellappa, D.W. Jacobs. (2016). Frontal to Profile Face Verification in the Wild, IEEE Conference on Applications of Computer Vision, 2016.
36. Wang, Zhongyuan, Baojin Huang, Guangcheng Wang, Peng Yi, and Kui Jiang. "Masked face recognition dataset and application." IEEE Transactions on Biometrics, Behavior, and Identity Science (2023).
37. Gao, P., Wu, W., & Li, J. (2021). Multi-source fast transfer learning algorithm based on support vector machine. Applied Intelligence, 1-15.
38. K. He, X. Zhang, S. Ren and J. Sun. (2016). Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016), pp. 770-778, doi: 10.1109/CVPR.2016.90.
39. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna. (2016). Rethinking the Inception Architecture for Computer Vision. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016), pp. 2818-2826, doi: 10.1109/CVPR.2016.308.
40. G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger. (2017). Densely Connected Convolutional Networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), pp. 2261-2269, doi: 10.1109/CVPR.2017.243.
41. K. Simonyan, A. Zisserman. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. Computing Research Repository (CoRR) - arXiv eprint arXiv: /abs/1409.1556.

42. Viola, P.; Jones, M. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 8–14 December 2001; pp. 511–518.

43. Benedict, S. R., & Kumar, J. S. (2016, October). Geometric shaped facial feature extraction for face recognition. In 2016 IEEE International Conference on Advances in Computer Applications (ICACA) (pp. 275-278). IEEE.

44. Trockman, A., & Kolter, J. Z. (2022). Patches Are All You Need?. arXiv preprint arXiv:2201.09792.

45. Shaheed, Kashif, Aihua Mao, Imran Qureshi, Qaisar Abbas, Munish Kumar, and Xingming Zhang. "Finger-vein presentation attack detection using depthwise separable convolution neural network." Expert Systems with Applications 198 (2022): 116786.

46. Thilagavathi, B., Suthendran, K., & Srujanraju, K. (2021). Evaluating the AdaBoost Algorithm for Biometric-Based Face Recognition. In Data Engineering and Communication Technology (pp. 669-678). Springer, Singapore.