Advanced Masked Face Recognition using Robust and Light Weight Deep Learning Model

Md. Omar Faruque
Hajee Mohammad Danesh Science and Technology University
Computer Science and Engineering
Dinajpur, Bangladesh

Md. Rashedul Islam
Hajee Mohammad Danesh Science and Technology University
Computer Science and Engineering
Dinajpur, Bangladesh

Md. Touhid Islam
Hajee Mohammad Danesh Science and Technology University
Computer Science and Engineering
Dinajpur, Bangladesh

ABSTRACT
For public health and safety reasons, face masks were required worldwide during the COVID-19 epidemic. However, this poses challenges for face recognition systems as the face is partially covered. Face recognition is a widely used and cost-effective biometric security system, but it faces difficulties in accurately identifying individuals wearing masks. Existing algorithms for face recognition have struggled to maintain efficiency, accuracy, and performance in the context of masked faces. To address these challenges and improve cost-effectiveness, a new machine learning model is required. This manuscript describes a lightweight deep learning methodology that is flexible and efficient in recognizing masked faces. The HSTU Masked Face Dataset (HMFD) is utilized, comprising frontal and lateral faces with various colored masks. Our proposed method involves a lightweight CNN model designed to enhance the accuracy of masked face identification. To enhance operational efficiency, methods like batch normalization, dropout, and depth-wise normalization are integrated which are tailored to meet particular specifications, aiming to optimize overall performance. These techniques improve the efficiency and accuracy of the model while minimizing overall complexity. In this research, the accuracy of the model is evaluated in comparison to other well-established deep learning models, including VGG16, VGG19, Extended VGG19, MobileNet, and MobileNetV2. The results demonstrate that our lightweight deep learning model outperforms these models, achieving a high recognition accuracy of 97%. By considering the needs of the task and carefully optimizing the model architecture, our proposed method offers an effective and efficient solution for recognizing masked faces in real-world scenarios.

General Terms
Graphics and Imaging, Computer Vision, Pattern Recognition et. al.

Keywords
Masked Face Recognition, Deep Learning, Convolutional Neural Network, Max-pooling, Lightweight CNN, Covid-19 Pandemic

1. INTRODUCTION
There has been a lot of interest in face recognition technology in the past several decades, and researchers all over the world have been working hard to perfect the technology [1-5]. Considerable gains have been achieved in this domain through the development of technology and artificial intelligence [6-7]. Consequently, there has been a widespread use of facial recognition systems by both public and commercial organizations with the aim of bolstering security measures and managing access control across diverse environments such as airports, educational institutions, workplaces, and communal areas [8-12]. The COVID-19 pandemic has prompted government authorities to enforce biosafety measures in order to mitigate the transmission of illnesses [13-15]. One rule that has been implemented is the compulsory utilization of facial coverings in public settings, as empirical evidence has demonstrated their efficacy in safeguarding persons and the surrounding population [16-18]. Nevertheless, the implementation of this fundamental procedure has presented security obstacles in contemporary technology. Security systems that rely on face recognition technology encounter challenges when individuals wear masks, which is a common occurrence in various public settings like transportation hubs, shopping malls, schools, and companies [19]. Traditional face identification systems often rely on persons fully exposing their faces to the camera in order to accurately recognize and identify them. However, this approach becomes unworkable in light of the current advise to wear masks when outside the home [20]. In order to effectively mitigate this concern and bolster security protocols, the present study posits the use of a concealed facial recognition framework. This research study has made several significant contributions, which are outlined below:

1. The proposed model is a lightweight deep learning model specifically designed for masked face recognition, taking into account the reduced Region of Interest (ROI) caused by face masks.
2. Our model employs a minimal number of layers and incorporates techniques such as dropout, batch normalization, and depthwise batch normalization to improve performance.
3. A comparative analysis of the proposed model is conducted with well-known deep learning models, including VGG16, VGG19, Mobilenet, MobilenetV2, and Extended VGG19. The effectiveness and efficiency of masked face recognition are appraised by the proposed model through the evaluation of outcomes derived from these models.
4. This study primarily aims to overcome the issues presented by face masks in image recognition systems through the introduction of a lightweight Convolutional Neural Network (CNN) model. Additionally, a thorough comparison is conducted between this proposed model and other well-established models.

2. LITERATURE REVIEW
In light of the prevailing COVID-19 circumstances, the utilization of face masks has been deemed obligatory as a means to mitigate the transmission of the virus. Nevertheless, this is a considerable obstacle for face recognition systems, as

they significantly depend on facial characteristics to achieve precise identification. The objective of this study, as described in citation [21], is to overcome the constraints encountered by facial recognition systems in handling masked persons by the utilization of sophisticated deep learning techniques. This work suggests the utilization of two distinct datasets, namely the Masked Face Detection Dataset (MFDD) and the Real-World Masked Face Recognition Dataset (RMFRD), as a means to address the difficulties associated with identifying faces that are covered by masks. This approach is motivated by the absence of publicly accessible datasets that are especially tailored for this particular objective.

Furthermore, this study presents a revolutionary methodology known as RetinaFaceMask, which incorporates a face mask detector that is both extremely precise and efficient. The purpose of this strategy is to tackle the public health challenges that have emerged as a result of the COVID-19 epidemic. The RetinaFaceMask system utilizes a feature pyramid network to integrate high-level semantic information from many feature maps. Additionally, a novel context attention module has been incorporated to enhance the precision of face mask identification. The approach outlined in this research showcases a heightened degree of precision, exhibiting enhancements of 2.3% and 1.5% in comparison to the presently existing state-of-the-art procedures. These advancements were assessed utilizing a publically accessible dataset that was purposefully created for the purpose of detecting face masks. In addition, it is worth noting that the RetinaFaceMask model demonstrates recall improvements of 11.0% and 5.9% in comparison to the baseline model as reported in reference [22]. The main aim of this study is to investigate the application of Non-Negative Sparse Coding (NNSC) as a technique for extracting features in the field of face recognition. The research also encompasses a comparative examination of NNSC in relation to other part-based methodologies, specifically Non-negative Matrix Factorization (NMF) and Local-Non-negative Matrix Factorization (LNMF).

The NNSC approach has undergone evaluation on many databases, such as the Aleix-Robert (AR), Face Recognition Technology (FERET), Yale B, and Cambridge ORL databases [23]. Furthermore, this research examines the efficacy of artificial neural networks (ANN) in the domain of image processing and pattern identification, specifically focusing on facial recognition within this discipline. Artificial Neural Networks (ANN) have demonstrated significant efficiency in comparison to alternative methodologies [24]. Additionally, this study investigates the utilization of restricted storage capacity in face authentication by the implementation of compressed facial pictures, while simultaneously ensuring the preservation of elevated recognition rates. Furthermore, the study included an examination of several Regions of Interest (ROI) masks and their corresponding rates of facial recognition [25].

The primary objective of this study is to examine the inherent difficulties and intricacies associated with facial recognition systems, which arise from the wide range of traits seen in human faces. These features comprise a range of parameters including color, expression, location, posture, and orientation. In order to discern various facial expressions, a composite of modeling approaches is employed, which considers the motions of the eyes, mouth, nose, and other face attributes (26). Furthermore, the study presents a novel co-mining methodology that has the ability to train datasets including diverse degrees of noise. Loss values are applied as markers to identify noisy labels, while clean face photos with high confidence are used to address issues that arise from sample-selection bias. The efficacy of this methodology is exemplified by conducting comparisons with cutting-edge alternatives, employing three extensively utilized datasets, and evaluating performance across several benchmarks [27]. Moreover, the present work centers on a benchmark assignment that entails the recognition of a vast number of celebrities by analyzing face images. Subsequently, the study aims to establish the connection between these persons and their respective entity keys. Disambiguation techniques are utilized to improve the accuracy of recognition, therefore providing notable contributions to many practical applications, notably within the realm of knowledge base systems. In order to get the most favorable outcomes, the researchers engage in the endeavor of formulating and delivering precise sets of measurements, an assessment technique, and a dataset for training purposes [28].

The authors of the present study [24] propose a novel model known as DeepMaskNet, which demonstrates efficacy in the detection of face masks and subsequent recognition of masked faces. The authors utilize a method of experimentation, employing stochastic gradient descent (SGD) as the training technique for the DeepMaskNet model. Furthermore, a scholarly investigation carried out by [19] provides a comparative evaluation of the VGG16 and MobileNetV2 models in the specific domain of identifying faces that are obscured by masks. The findings of the study suggest that MobileNetV2 exhibits better performance than VGG16, demonstrating higher performance within the particular area under investigation. Nevertheless, it is crucial to acknowledge that a significant portion of the current body of research in this particular domain heavily depends on enhanced datasets that are drawn from pre-existing datasets originally designed for the purpose of face recognition. The objective of this research is to develop a robust facial recognition system specifically designed for faces wearing masks, focusing on the utilization of the HSTU Masked Face Dataset (HMFD) [20].

The authors then describe a deep learning strategy that makes effective use of available resources to accomplish the task of accurate facial recognition even when subjects are concealed behind masks. The primary aim of this study is to improve the reliability and effectiveness of facial recognition algorithms that are designed to identify individuals wearing masks. This will be achieved by utilizing a dataset consisting of actual images of individuals wearing masks. The proposed lightweight deep learning approach greatly enhances the precision of facial recognition, particularly when faced with the difficulty of identifying persons wearing masks. The findings of this research provide a significant contribution to the development of effective and efficient techniques for facial identification in real-world situations characterized by the prevalent use of masks.

3. METHODOLOGY

3.1 Convolutional Neural Network

This research paper introduces a system for the identification of obscured facial features, encompassing several stages such as the creation of a model, the process of training, conducting tests, and evaluating performance. The objective of the proposed methodology is to tackle the difficulties related to facial recognition in the presence of masks. Fig. 1 provides an overview of the process, visually representing the sequential progression of the many stages involved. Section III of the paper discusses the HSTU Masked Face Dataset (HMFD) and provides a detailed description of the dataset. The authors

explain how the dataset was obtained and highlight the main preprocessing procedures that were performed to ensure its suitability for training machine learning models.

CNNs have been developed with the purpose of capturing and extracting crucial information from input data, namely within the domain of pictures. This functionality allows humans to execute a range of computer vision activities, encompassing but not limited to tasks such as picture classification, object identification, and image segmentation. CNNs are able to do this task by efficiently discerning spatial hierarchies and patterns within the given data. This is made possible by the incorporation of specialized layers such as convolutional layers, pooling layers, and fully connected layers. This methodology guarantees a comprehensive and accurate examination.

In contrast, transfer learning entails using the capabilities of pre-existing deep neural network architectures such as VGG16, VGG19, MobileNet, and MobileNetV2. The models are equipped with acquired weights and feature representations obtained through rigorous training on big datasets. The utilization of transfer learning allows for the acceleration and enhancement of new machine learning tasks through the application of pre-trained models. One of the primary benefits of transfer learning is its capacity to utilize the acquired features from pre-trained models, enabling effective training despite having a limited amount of data. The utilization of this technique facilitates the efficient transfer of knowledge from jobs that possess abundant data to those that possess limited datasets, resulting in significant savings in terms of time and resources. This methodology is especially advantageous in situations where there is a scarcity of data, since it aids in addressing the difficulty of training deep neural networks from the beginning.

The expanded VGG19 model is a modification of the original VGG19 architecture. The model incorporates supplementary components, namely a Dense layer of 4096 nodes, a Dropout layer with a dropout rate of 50%, and a Batch Normalization layer. The inclusion of additional layers in the neural network serves the purpose of mitigating overfitting, enhancing model stability, and standardizing and normalizing the network. The extended VGG19 architecture demonstrates high efficacy when applied to the dataset pertaining to masked face recognition. The authors employ grid search methodology to select the hyperparameters for optimizing the expanded VGG19 model. The hyperparameters under consideration encompass activation functions, optimizer selection, dropout rate determination, and the specification of the number of nodes in the dense layer. The empirical findings and research indicate that the Adam optimizer exhibits strong performance in the context of classification problems. Additionally, it is widely accepted in the field to employ a dropout rate of 50% as a typical strategy for mitigating overfitting.

The 2D operation of a CNN can be represented as follows:

$$y = f(W \times x + b) \qquad (1)$$

The variable 'x' is commonly used to denote the input data, typically in the form of a two-dimensional image. The letter 'W' is used to represent the learnable parameters in a neural network, specifically referring to the convolutional filters or kernels. The variable 'b' denotes the biases that correlate to each filter. The symbol 'f' is used to represent the activation function, which serves the purpose of introducing non-linearity into the output. Ultimately, the variable 'y' denotes the resultant output feature map subsequent to the convolutional procedure. The aforementioned parts collectively comprise the essential constituents in CNNs for the purpose of image processing jobs. Multiple convolutional layers are the backbone of the CNN design, which also commonly includes activation functions, pooling layers, and fully linked layers. The aforementioned layers, in conjunction with the parameters (W and b), are acquired through the training procedure in order to extract pertinent features and generate predictions based on the provided input data.

In this equation:

x represents the input data, which is typically a 2D image.
W denotes the learnable parameters known as the convolutional filters or kernels.
b represents the biases associated with each filter.
f refers to the activation function, which introduces non-linearity to the output.
y represents the output feature map after the convolutional operation.

The CNN architecture typically consists of multiple Multiple convolutional layers are the backbone of the CNN design, which also commonly includes activation functions, pooling layers, and fully linked layers. The layers, in conjunction with the parameters (W and b), are acquired through the training procedure to extract pertinent features and generate predictions on the provided input data. Max pooling is a frequently employed procedure within CNNs for the purpose of downsampling feature maps. The mathematical expression representing the max pooling procedure is given by:

$$y = maxpool(x, k, s) \qquad (2)$$

The variable 'x' denotes the input map of features. The variable 'k' represents the size of the pooling window, also known as the kernel, which is applied to the incoming data by a sliding mechanism. The variable 's' denotes the stride, which governs the magnitude of the step taken by the pooling window during movement. The precise mathematical expression for max pooling entails the process of finding the highest value within the pooling window.

$$y[i,j,c] = \max(x[i \times s : i \times s + k, j \times s : j \times s + k, c]) \qquad (3)$$

The formula denotes the output value at location (i, j) of the c-th feature map in the output tensor, represented by 'y[i, j, c]'. The input feature map x is subjected to a pooling window of size k × k. This pooling window is applied to the appropriate region of x, starting at position (i × s, j × s) and moving with a stride of s. The spatial dimensions of the feature maps are lowered through the repeated application of the max pooling operation, using suitable kernel size and stride. This allows the network to prioritize more significant features, leading to enhanced computational efficiency.

Batch normalization is a widely employed technique in deep neural networks that serves the purpose of normalizing the activations of intermediary layers. This intervention contributes to enhancing the network's stability and accelerating the training process. The mathematical expression for batch normalization can be formulated as follows:

$$y = \gamma \times \frac{(x - \mu)}{\sigma + \beta} \qquad (5)$$

In this equation:

'x' represents the input tensor or activations of a specific layer,

'γ ' denotes the learned scale parameter,
‘μ' represents the mean of the batch,
‘σ' represents the standard deviation of the batch,
‘β' denotes the learned shift parameter and .
'y represents the output after applying batch normalization.

The utilization of batch normalization has been found to enhance the performance of neural networks through the mitigation of internal covariate shifts, facilitating the use of greater learning rates, and imparting a regularization effect. The inclusion of normalization techniques in the training process aids in the stabilization and acceleration of the network's training by guaranteeing that the inputs of the network are maintained within a normalized and consistent range across the various layers.

One variant of batch normalization, known as "depthwise batch normalization," does normalize independently on each channel of a depthwise convolutional layer. Normalizing the activations inside each channel aids in the normalization of the learning process and facilitates the convergence of the neural network. The mathematical expression for depthwise batch normalization is as follows:

$$y[c] = \gamma[c] \times \frac{(x[c] - \mu[c])}{\sigma[c] + \beta[c]} \quad (6)$$

In this equation:

x[c] represents the input activations for the c-th channel.,

γ[c] denotes the learned scale parameter specific to the c-th channel, .

μ[c] represents the mean of the activations in the c-th channel, .

σ[c] represents the standard deviation of the activations in the c-th channel, .

β[c] denotes the learned shift parameter specific to the c-th channel and .

y[c] represents the output activations of the c-th channel after applying depthwise batch normalization.

By applying batch normalization, the network can benefit from improved learning dynamics and convergence within each channel independently. It helps to normalize the activations within each channel, ensuring that the network can learn effectively from the specific features present in each channel.

CNNs make use of a particular convolutional layer called the Depthwise Convolutional Layer. It applies a unique convolutional filter to each input channel, thus each channel may be processed separately. That's why the feature maps this layer outputs have the same amount of channels as the input.

The equation for the depthwise convolution operation can be represented as follows:

$$Y[:,:,k] = \sum(W[:,:,k] \times X[:,:,k]) \quad (7)$$

In this equation:

Y[:,:,k] represents the k-th output feature map, .
W[:,:,k] denotes the k-th convolutional filter/kernel specific to the k-th channel, .
X[:,:,k] represents the k-th input channel.

For each output feature map, the depthwise convolution operation convolves the corresponding filter W[:,:,k] with the input channel X[:,:,k] using element-wise multiplication and summation. The resulting output feature map Y[:,:,k] captures the spatial correlations and patterns within that particular channel.

3.2 Proposed Light Weight CNN Model

The suggested Lightweight CNN model is an augmentation of the original CNN architecture, wherein Batch Normalization is incorporated in each layer prior to the input being passed to the fully connected layer. Fig. 1 depicts the comprehensive architecture of the lightweight variant of the conventional CNN model, which demonstrates notable efficacy when applied to our specific dataset.

Table 1. Layered exhaustive summary of the proposed Lightweight CNN model with 264 classes of utilized dataset.

Layer (type)	**Output Shapes**	**Params #**
Conv2D	(None, 146, 146, 64)	4864
MaxPooling2D	(None, 73, 73, 64)	0
Conv2D	(None, 69, 69, 128)	204928
MaxPooling2D	(None, 34, 34, 128)	0
Conv2D	(None, 32, 32, 256)	295168
MaxPooling2D	(None, 16, 16, 256)	0
Depthwise Conv 2D	(None, 14, 14, 256)	2560
MaxPooling2D	(None, 7, 7, 256)	0
Flatten	(None, 12544)	0
Dense	(None, 4096)	51384320
Dropout	(None, 4096)	0
Dense	(None, 264)	1081608
Total params: 52,976,264, Trainable params: 52,974,856,		

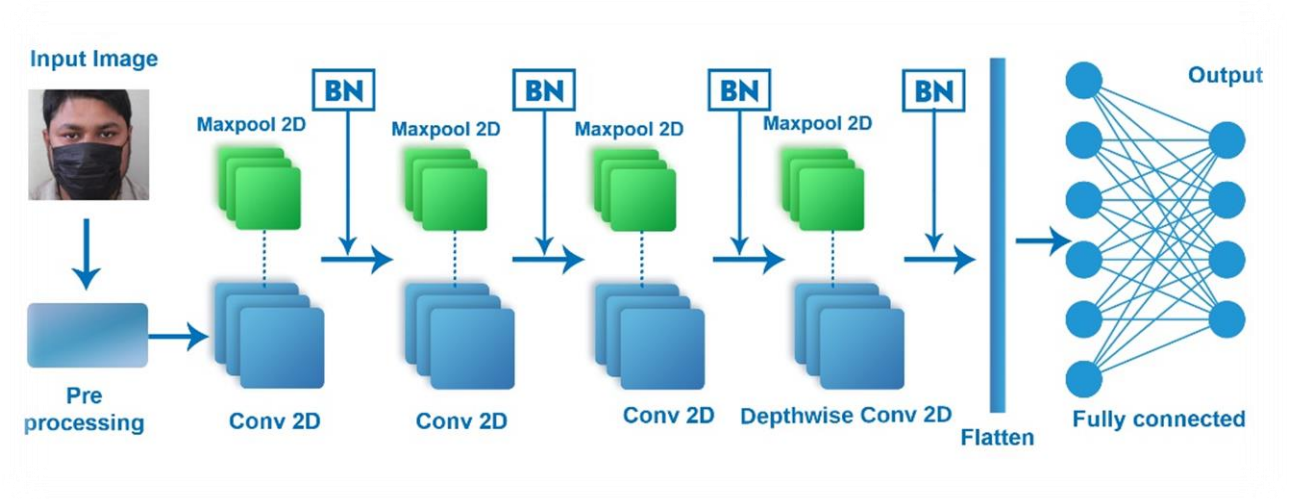

Fig. 1. Flow diagram of the proposed masked face recognition model

The present study presents a newly developed CNN structure specifically designed to enhance the accuracy of facial recognition in situations where individuals are wearing masks. The present design effectively capitalizes on the benefits of depthwise separable convolutions and incorporates many layers to attain an ideal equilibrium between computing efficiency and accuracy. The CNN model demonstrates exceptional performance in scenarios when face characteristics are partially obscured by masks, rendering it very pertinent within the context of the prevailing COVID-19 epidemic. This novel methodology guarantees the efficient and precise identification of faces by the model, even in cases when some facial characteristics are obscured.

The CNN model is constructed with the Keras package in Python, with an emphasis on maintaining a minimalistic design. The process starts with the establishment of an initial sequential model. Next, a sequence of convolutional layers is introduced. The initial layer of this sequence employs 64 filters and a kernel size of (5,5), accompanied by a Rectified Linear Unit (ReLU) activation function. After the first layers, a max-pooling layer is incorporated, followed by the inclusion of a batch normalization layer. This measure is used in order to augment the model's ability to acquire knowledge and apply it to new situations.

Subsequently, a secondary convolutional layer is incorporated, with 128 filters and a kernel size of (5,5), while employing the ReLU activation function once more. Subsequently, an additional max pooling layer and batch normalization layer are incorporated to enhance the performance of the model. The architectural design proceeds by incorporating a subsequent convolutional layer, characterized by the utilization of 256 filters and a kernel size of (3,3), along with the application of the Rectified Linear Unit (ReLU) activation function. In addition, the architecture incorporates a max pooling layer and a batch normalization layer.

A depthwise convolutional layer with a (3,3) kernel size and ReLU activation is included to induce non-linearity and lower the model's complexity. Following that, an additional layer of max pooling and batch normalization is incorporated. Subsequently, the resulting output is compressed into a single-dimensional feature vector using flattening the multi-dimensional feature maps. To enable the representation of more intricate features, a fully linked layer of 4096 nodes employing the Rectified Linear Unit (ReLU) activation function is incorporated. Detailed parameters of the model is presented in Table 1. Furthermore, the model incorporates a dropout layer with a dropout rate of 0.5 to mitigate the issue of overfitting and enhance the model's ability to generalize. In conclusion, a fully connected layer consisting of 264 nodes and employing softmax activation is appended as the last layer to categorize the masked faces into their respective classes.

As a loss function, Sparse Categorical Cross entropy is ideal for multi-class classification applications like face recognition, and it is used in the model's compilation. The Adam optimizer is utilized to effectively update the parameters of the model throughout the training process, while the accuracy metric is selected as the criterion to assess the performance of the model. In summary, the lightweight CNN model approach introduces a novel CNN architecture that is specifically designed to identify faces that are covered by masks. Some of the model's success can be attributed to the use of depthwise separable convolutions, along with other carefully chosen layers.

4. RESULT AND DISCUSSIONS

4.1 Dataset Description

The objective of this study is to overcome the shortcomings of current masked face datasets by introducing a new dataset known as the HSTU Masked Face Dataset (HMFD) [29]. This dataset is designed to enhance the precision of face identification while individuals are wearing masks. Image augmentation techniques have been used to build a number of datasets, such as the conversion of Labeled Faces in the Wild (LFW) [30] and the CASIA face anti-spoofing [31], however these datasets do not accurately portray the appearance of masked faces [32]. Within these augmented datasets, it is common for the cheeks to remain uncovered, so deviating from real-world circumstances when masks typically cover the mouth, nose, and cheeks. In order to address these constraints, the novel dataset HMFD was developed, which incorporates photos of individuals who are appropriately wearing masks, ensuring that the mouth, nose, and cheeks are totally covered. The dataset includes a wide variety of perspectives, ages, and genders. The Dataset has multiple samples, each including a collection of 20 photos. These images encompass a variety of facial representations, including faces with masks of different colors (blue and black), faces without masks, and faces with masks that are not worn properly. The dataset known as HMFD comprises a total of 5,280 pictures, which have been collected from a sample of 264 individuals.

The dataset has been prepared for training in machine learning, encompassing essential preprocessing activities such as foldering, labeling, renaming, segmentation, and scaling. The raw images undergo a segmentation process in order to isolate and extract the specific region of interest (ROI). Following this, every image is shrunk to dimensions of 150×150 pixels in order to decrease the overall data size and improve the effectiveness of the training process.

The High-quality Masked Face Dataset (HMFD) serves as a reliable dataset that faithfully captures the visual characteristics of masked faces, facilitating the advancement and assessment of resilient masked face recognition algorithms. This study makes a valuable contribution to the advancement of masked face recognition in the field, specifically addressing the difficulties encountered in real-world situations where individuals wear masks. By doing so, it aids in the enhancement of security and identity systems across several domains.

4.2 Experiment Setup

The present work introduces a methodology for the identification of obscured facial features, encompassing a series of sequential stages such as the formulation of a model, the process of training, the execution of testing, and the assessment of performance. The objective of the suggested methodology is to effectively tackle the difficulties linked to the identification of individuals when they are wearing facial masks. The technique overview is presented in Fig. 1, which visually represents the sequential progression of the many stages involved. The procedure of dataset preprocessing is outlined in Section III. The collection and preprocessing of the HSTU Masked Face Dataset (HMFD) are carried out to ensure its suitability for training machine learning models. Extensive tests were done to assess the performance of the proposed approach. The dataset utilized for training and evaluating the models was the HMFD dataset, which had 5280 photos featuring 264 persons wearing masks. With an 80:20 split, a subset of the data for training and another for testing is used.

4.3 Performance Measure Metrics

In order to assess the efficacy of the models in detecting obscured facial features, a range of evaluation criteria are employed, encompassing accuracy, precision, recall, and F1-score.

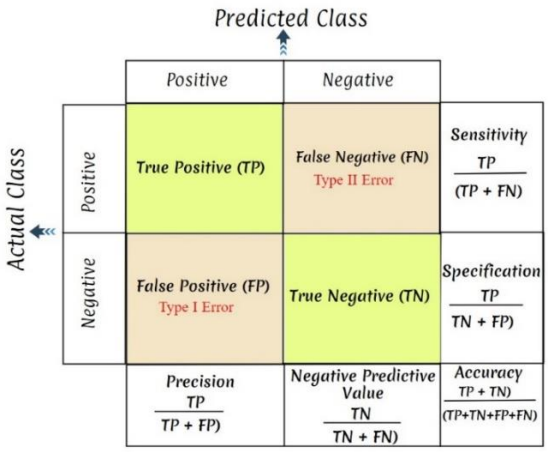

Fig. 2. Brief summary of a confusion Matrix

In the context of binary classification, the term "True Positives" (TP) refers to the right identification of positive samples, whereas "True Negatives" (TN) denotes the proper classification of negative data. On the other hand, "False Positives" (FLP) refer to the incorrect identification of negative samples as positive, while "False Negatives" (FLN) represent the inaccurate labeling of positive samples as negative. The aforementioned criteria are of utmost importance for assessing the effectiveness of categorization algorithms. A concise overview of a confusion matrix and the diverse performance metrics presented in Fig. 2.

4.3.1. Overall Accuracy

The measurement of overall accuracy is a crucial performance indicator that quantifies the ratio of correctly categorized data examples to the overall number of data occurrences. The assessment offered is a thorough review of the effectiveness of the classification model in effectively identifying hidden face characteristics.

The accuracy of a classification model may be determined by computing the ratio of properly classified samples to the total number of samples. Following this, the equation that represents accuracy may be expressed as:

$$\text{Accuracy} = \frac{(\text{TRP} + \text{TrN})}{(\text{TRP} + \text{TRN} + \text{FLP} + \text{FLN})} \quad (8)$$

In regard to a binary classification task, accuracy is a metric that quantifies the ratio of accurately categorized samples to the overall number of samples. This metric is particularly relevant when dealing with two distinct classes, such as masked and unmasked faces.

4.3.2. Precision

Precision is a quantitative measure that assesses the ability of a model to provide accurate positive predictions. The measure refers to the proportion of correctly predicted positive cases divided by the total number of instances anticipated as positive. A high accuracy score suggests that the model has a comparatively lower incidence of false positives. The mathematical representation denoting accuracy can be formally defined as:

$$\text{Precision} = \frac{\text{TRP}}{(\text{TRP} + \text{FLP})} \quad (9)$$

The precision indicates the closely value to the overall accuracy of a model.

4.3.3. Recall

The metric of recall, also known as sensitivity or true positive rate, assesses the model's capacity to accurately predict positive occurrences. The statistic measures the proportion of correctly predicted positive cases in relation to the overall number of positive instances. A high recall score indicates that the model exhibits a reduced frequency of false negatives. The mathematical representation of recall, also known as sensitivity or true positive rate, may be formally described as follows:

$$\text{Recall} = \frac{\text{TRP}}{(\text{TRP} + \text{FLN})} \quad (10)$$

4.3.4. F1 Score

The F1-score is a balanced indication since it gives equal weight to both the accuracy and the recall of the information being evaluated. When there is a disparity in the number of positive and negative classifications present in the dataset,

which often occurs in real-world scenarios, the applicability of this method becomes readily apparent. The following is a definition of the mathematical expression that may be used to describe the F1 score:

$$\text{F1 Score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \qquad (11)$$

Furthermore, the evaluation of validation features, such as validation loss and a validation accuracy curve, is also taken into account. The validation accuracy curve depicts the fluctuations in the model's performance throughout the training phase in a visual format. Consequently, it offers valuable insights about the course of learning. On the other hand, the validation loss serves as an indicator of the model's ability to generalize to unfamiliar and unseen data. lesser values of the validation loss suggest a greater degree of generalization, whereas larger values imply a lesser level of generalization. The objective of this study is to assess the efficacy of various models in accurately identifying veiled faces, as well as to ascertain the practicality of the proposed approach in addressing the difficulties posed by facial masks. The achievement of this objective will be facilitated by analyzing the aforementioned metrics.

4.4 Result Analysis

The results of the initial phase of our tests are presented in Table 2. The models underwent training using the designated training set and were subsequently assessed based on their performance on the designated testing set. The performance was assessed by employing criteria like as accuracy, precision, recall, and F1-score. The empirical findings of our proposed lightweight CNN model exhibit its comparative advantage over alternative models concerning the recognition of masked faces. The model demonstrates notable performance measures, encompassing a testing accuracy of 96%, precision of 97%, recall of 96%, and F1 score of 96%. These measures serve as indicators of the model's capacity to effectively classify faces that have been obscured. In our study, we observed that the pre-trained CNN, SVM, VGG16, VGG19, Mobilenet, and MobilenetV2 models yielded a testing accuracy of 87% when compared. The diminished level of accuracy seen can be ascribed to the model's inadequate training on our particular dataset. In order to mitigate this constraint and enhance the efficiency, we implemented alterations in our suggested lightweight CNN. The proposed alterations entail the incorporation of a supplementary dense layer featuring a dropout rate of 50%, a batch normalization layer, and the integration of depthwise convolution in conjunction with the pre-trained CNN model.

Table 2. Masked Face Recognition Performance Evaluation (Frontal Image as Testing Samples)

Model Name	Accuracy	Precision	Recall	F1 Score
VGG16	87.00	88.00	87.00	85.00
VGG19	74.00	86.00	86.00	84.00
Extended VGG19	96.00	97.00	96.00	96.00
Mobilenet	60.00	61.00	60.00	56.00
MobilenetV2	60.00	59.00	60.00	56.00
Lightweight CNN	98.00	98.00	97.00	98.00

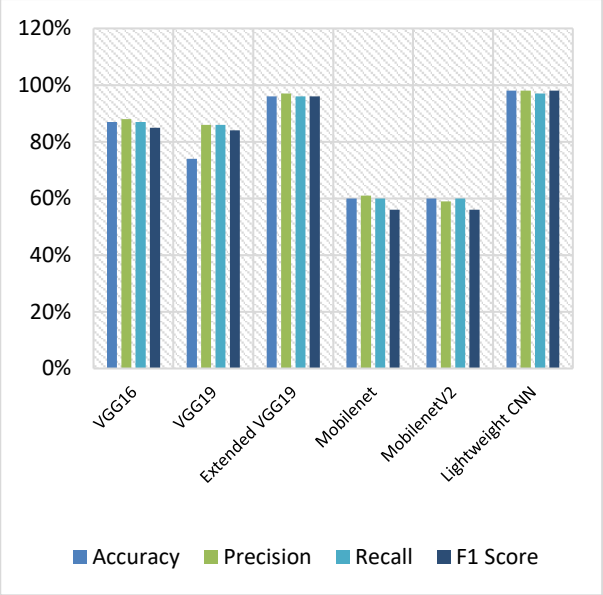

Fig. 3. Masked Face Recognition Performance Evaluation with Bar Chart (Frontal Image as Testing Samples)

Table 3. Masked Face Recognition Performance Evaluation (Lateral Image as Testing Samples)

Model Name	Accuracy	Precision	Recall	F1 Score
VGG16	65.00	77.00	65.00	65.00
VGG19	64.00	7.00	64.00	65.00
Extended VGG19	77.00	80.00	77.00	76.00
Mobilenet	33.00	41.00	33.00	31.00
MobilenetV2	34.00	44.00	34.00	33.00
Lightweight CNN	79.00	83.00	80.00	79.00

Due to these improvements, the lightweight CNN demonstrated a notable increase in accuracy, reaching 96%. This surpassed the performance of the conventional CNN model. Furthermore, the precision, recall, and F1 score exhibited remarkable values of 97%, 96%, and 96% correspondingly, so showcasing the model's proficiency in effectively categorizing obscured faces. The performance of each model is represented graphically in Fig. 3 by the validation accuracy of each model. In comparison to both pre-trained models and conventional CNN architectures, our experimental results show that our proposed lightweight CNN model is significantly more effective. Fig. 2 depicts the validation loss of the models, confirming the usefulness of our proposed lightweight CNN in the context of masked face recognition. In a similar vein, Fig. 4 depicts the validation loss of the models.

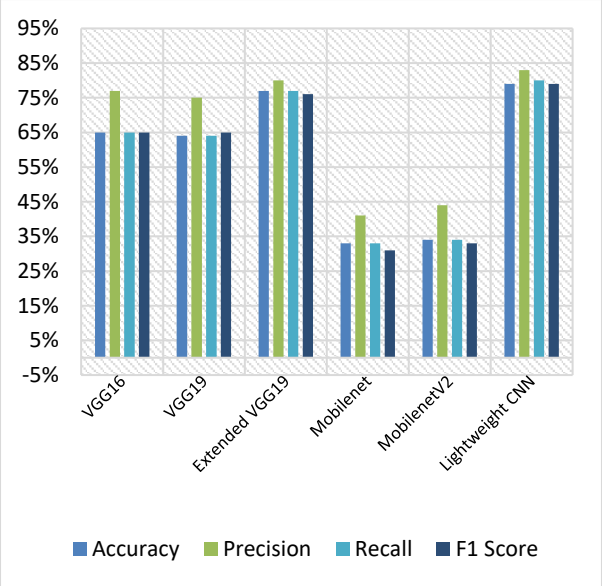

Fig. 4. Masked Face Recognition Performance Evaluation with Bar Chart (Lateral Image as Testing Samples)

The findings of the study indicate that the Lightweight CNN model we suggested had superior performance compared to other standard deep learning methods. The model demonstrated a testing accuracy of 96%, suggesting its efficacy in facial recognition tasks involving individuals wearing masks. The criteria of precision, recall, and F1-score also demonstrated favorable outcomes, so providing additional confirmation of the effectiveness of our methodology.

In the second part of our investigation, we used side views as test photos to see how well each model performed. The data obtained in this phase exhibit variations when compared to the frontal photographs. Table 3 presents a comprehensive summary of the performance measures attained by each model.

When lateral images were used for testing, the standard CNN model demonstrated an accuracy of 60%, precision of 63%, recall of 60%, and an F1 score of 59%. The VGG19 model exhibited marginally superior performance, attaining an accuracy of 61%, precision of 66%, recall of 61%, and an F1 score of 60%. Nevertheless, our proposed lightweight CNN model shown superior performance compared to CNN, support vector machine (SVM), VGG16, VGG19, Mobilenet, and MobilenetV2. When the lightweight CNN was tested with lateral images, it achieved an accuracy of 77%, precision of 80%, recall of 77%, and F1 score of 76% by including an additional dense layer with a 50% dropout layer, batch normalization, and depthwise convolution. It is imperative to acknowledge that the aforementioned findings exhibit relatively diminished performance in comparison to the outcomes derived from frontal images. This discrepancy suggests the heightened challenge associated with the identification of masked faces when observed from lateral perspectives.

The validation accuracy curves presented in Fig. 3 provide additional insight into the performance of the three models when tested on both frontal and lateral images. The lightweight CNN model consistently demonstrates superior performance compared to the other models across all scenarios. The visual representation of lateral pictures may exhibit a relatively diminished strength in terms of accuracy when compared to frontal images, although this observation does not necessarily indicate a substantial decline in overall performance. The lower accuracy threshold of the lightweight CNN model for lateral pictures is the minimal attainable level of accuracy in crucial scenarios. Furthermore, Fig. 4 illustrates the validation loss, indicating a strong alignment between the dataset and the models. The loss exhibits a declining trend as the number of epochs increases. The lightweight CNN model demonstrates higher performance in comparison to alternative approaches.

Although the performance of the suggested model demonstrates satisfactory results in the context of masked face recognition, it is crucial to acknowledge the influence of head position on the outcomes. The accuracy of the testing reduces when utilizing lateral images, which depict the most extreme head position or the highest angle of deviation from the frontal position. This phenomenon occurs due to the variability in facial information as a function of the face's angle. The lightweight CNN model has a 96% accuracy rate when applied to frontal images while achieving a 77% accuracy rate when applied to lateral images. Despite the comparatively modest percentage of 77%, this value signifies that our model consistently achieves a minimum accuracy of 77% even when faced with the most demanding or crucial head angles.

The enhanced efficacy of the lightweight CNN model can be ascribed to the incorporation of supplementary components, including the additional dense layer, batch normalization layer, and dropout layer. The inclusion of a dense layer in the model facilitates its ability to adapt to the dataset at hand, while simultaneously capitalizing on the knowledge encoded in the pre-trained weights. Batch normalization is a technique that standardizes the inputs of each layer in order to expedite and enhance the training process. The inclusion of a dropout layer with a dropout rate of 50% serves to mitigate the issue of overfitting and facilitates the generalization capabilities of the model.

In general, the lightweight CNN model described in this work exhibits superior performance in comparison to the other models employed. Although there is a decrease in accuracy when using lateral images, the model continually demonstrates strong performance, maintaining a minimum accuracy of 77% for important head angles. The incorporation of additional components such as a dense layer with increased complexity, a batch normalization layer for improved training stability, a depthwise convolution layer for enhanced feature extraction, and a dropout layer for regularization collectively enhance the performance and adaptability of the model in the context of masked face recognition tasks.

5. CONCLUSION AND FUTURE WORK

The primary emphasis of the essay centers around the development of a deep learning model specifically tailored for the purpose of facial recognition in the presence of facial masks. The deep learning model that has been developed is a variant of the CNN that is designed to be lightweight. Empirical findings from experiments conducted indicate that this model exhibits greater performance when compared to other traditional deep learning methodologies. A notable finding derived from the inquiry is that the proposed model demonstrates a heightened level of accuracy (reaching up to 97%) in the recognition of frontal masked faces, as opposed to lateral masked faces. This implies that the efficacy of the model may differ based on the angle or direction of the obscured facial features. The researchers want to pursue the development of a real-time masked face recognition system in their forthcoming endeavors, aiming to get precise outcomes through the use of a reduced training dataset. Furthermore, the researchers want to investigate facial expressions by enhancing the dataset with a

broader range of diverse and plentiful data, encompassing both increased amount and variation. In general, the paper emphasizes the potential of the suggested deep learning model for recognizing masked faces and provides suggestions for future enhancements and developments in this area of research.

6. REFERENCES

[1] Adjabi, I., Ouahabi, A., Benzaoui, A., & Taleb-Ahmed, A. (2020). Past, Present, and Future of Face Recognition: A Review. Electronics, 9(8), 1188. https://doi.org/10.3390/electronics9081188.

[2] Ko, B. (2018). A Brief Review of Facial Emotion Recognition Based on Visual Information. Sensors, 18(2), 401. https://doi.org/10.3390/s18020401.

[3] Chakraborty, B. K., Sarma, D., Bhuyan, M. K., & MacDorman, K. F. (2018). Review of constraints on vision-based gesture recognition for human–computer interaction. IET Computer Vision, 12(1), 3-15.

[4] Egger, M., Ley, M., & Hanke, S. (2019). Emotion recognition from physiological signal analysis: A review. Electronic Notes in Theoretical Computer Science, 343, 35-55.

[5] Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. Computational intelligence and neuroscience, 2018.

[6] Dang, K., & Sharma, S. (2017, January). Review and comparison of face detection algorithms. In 2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence (pp. 629-633). IEEE.

[7] Qiu, S., Liu, Q., Zhou, S., & Wu, C. (2019). Review of artificial intelligence adversarial attack and defense technologies. Applied Sciences, 9(5), 909.

[8] Galterio, M. G., Shavit, S. A., & Hayajneh, T. (2018). A review of facial biometrics security for smart devices. Computers, 7(3), 37.

[9] Cook, C. M., Howard, J. J., Sirotin, Y. B., Tipton, J. L., & Vemury, A. R. (2019). Demographic effects in facial recognition and their dependence on image acquisition: An evaluation of eleven commercial systems. IEEE Transactions on Biometrics, Behavior, and Identity Science, 1(1), 32-41.

[10] Jeon, B., Jeong, B., Jee, S., Huang, Y., Kim, Y., Park, G. H., ... & Choi, T. H. (2019). A facial recognition mobile app for patient safety and biometric identification: Design, development, and validation. JMIR mHealth and uHealth, 7(4), e11472.

[11] Dargan, S., & Kumar, M. (2020). A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities. Expert Systems with Applications, 143, 113114.

[12] Gonzalez-Sosa, E., Fierrez, J., Vera-Rodriguez, R., & Alonso-Fernandez, F. (2018). Facial soft biometrics for recognition in the wild: Recent works, annotation, and COTS evaluation. IEEE Transactions on Information Forensics and Security, 13(8), 2001-2014.

[13] Karthik, K., Babu, R. P. A., Dhama, K., Chitra, M. A., Kalaiselvi, G., Senthilkumar, T. M. A., & Raj, G. D. (2020). Biosafety concerns during the collection, transportation, and processing of COVID-19 samples for diagnosis. Archives of Medical Research, 51(7), 623-630.

[14] Ortiz, M. R., Grijalva, M. J., Turell, M. J., Waters, W. F., Montalvo, A. C., Mathias, D., ... & Leon, R. (2020). Biosafety at home: How to translate biomedical laboratory safety precautions for everyday use in the context of COVID-19. The American journal of tropical medicine and hygiene, 103(2), 838.

[15] Souza, T. M. L., & Morel, C. M. (2021). The COVID-19 pandemics and the relevance of biosafety facilities for metagenomics surveillance, structured disease prevention and control. Biosafety and Health, 3(01), 1-3.

[16] Mills, M., Rahal, C., & Akimova, E. (2020). Face masks and coverings for the general public: Behavioural knowledge, effectiveness of cloth coverings and public messaging. The Royal Society, 26.

[17] Liu, X., & Zhang, S. (2020). COVID-19: Face masks and human-to-human transmission. Influenza and other respiratory viruses, 14(4), 472.

[18] Ngan, M., Grother, P., & Hanaoka, K. (2020). Face recognition accuracy with masks using pre-COVID-19 algorithms. In NISTIR 8311.

[19] Zeng, J., Qiu, X., & Shi, S. (2021). Image processing effects on the deep face recognition system. Math. Biosci. Eng, 18(2), 1187-1200.

[20] Wang, Z., Huang, B., Wang, G., Yi, P., & Jiang, K. (2023). Masked face recognition dataset and application. IEEE Transactions on Biometrics, Behavior, and Identity Science.

[21] Wang, Z., Wang, G., Huang, B., Xiong, Z., Hong, Q., Wu, H., ... & Liang, J. (2020). Masked Face Recognition Dataset and Application (preprint).

[22] Fan, X., & Jiang, M. (2021, October). RetinaFaceMask: A single stage face mask detector for assisting control of the COVID-19 pandemic. In 2021 IEEE international conference on systems, man, and cybernetics (SMC) (pp. 832-837). IEEE.

[23] Shastri, B. J., & Levine, M. D. (2007). Face recognition using localized features based on non-negative sparse coding. Machine Vision and Applications, 18, 107-122.

[24] Kasar, M. M., Bhattacharyya, D., & Kim, T. H. (2016). Face recognition using neural network: a review. International Journal of Security and Its Applications, 10(3), 81-100.

[25] Fujita, M., Yoshida, T., & Hangai, S. (2006, May). A study on the effect of ROI masks on face recognition system using digital recorder. In 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings (Vol. 2, pp. II-II). IEEE.

[26] Saudagare, P. V., & Chaudhari, D. S. (2012). Facial expression recognition using neural network–An overview. International Journal of Soft Computing and Engineering (IJSCE), 2(1), 224-227.

[27] Guo, Y., Zhang, L., Hu, Y., He, X., & Gao, J. (2016). Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14 (pp. 87-102). Springer International Publishing.

[28] Wang, X., Wang, S., Wang, J., Shi, H., & Mei, T. (2019). Co-mining: Deep face recognition with noisy labels. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 9358-9367).

[29] Marjan, M. A., Hasan, M., Islam, M. Z., Uddin, M. P., & Afjal, M. I. (2022, December). Masked Face Recognition System using Extended VGG-19. In 2022 4th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE) (pp. 1-4). IEEE.

[30] Huang, G. B., & Learned-Miller, E. (2014). Labeled faces in the wild: Updates and new reporting procedures. Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep, 14(003).

[31] Boulkenafet, Z., Komulainen, J., & Hadid, A. (2015, September). Face anti-spoofing based on color texture analysis. In 2015 IEEE international conference on image processing (ICIP) (pp. 2636-2640). IEEE.

[32] Neto, P. C., Boutros, F., Pinto, J. R., Damer, N., Sequeira, A. F., & Cardoso, J. S. (2021, December). Focusface: Multi-task contrastive learning for masked face recognition. In 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021) (pp. 01-08). IEEE.