



A Novel Approach for Accurate Identification in Masked and Unmasked Scenarios using Glowworm Swarm Optimization and Neural Networks

Kosuri Naresh Babu¹ · Suneetha Manne²

Received: 15 December 2023 / Revised: 9 August 2024 / Accepted: 14 August 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Recognition and classification are among the most important applications of machine learning. This recognition process is used to identify objects and humans. In particular, it plays a major role in authentication processes by identifying features such as human eyes, fingerprints, and facial patterns. Among these features, facial recognition is an evolving technology used in smartphones, attendance systems in offices, and healthcare centers. Several research efforts have been conducted to perform facial recognition using machine learning and deep learning algorithms. These algorithms have performed well on faces without masks, but they have struggled with masked faces, as most facial features are hidden by the mask. Therefore, an improved algorithm is needed for performing facial recognition on faces with and without masks. Since the COVID-19 outbreak, research has been focused on using deep learning algorithms to identify masked faces. However, these algorithms were typically trained on faces both with and without masks. In this paper, we propose a facial recognition approach for recognizing faces with and without masks. The common regions of the face in both scenarios are identified by cropping the image. These cropped regions are then subjected to feature extraction using histogram properties, SURF, and SIFT features. The dominant features are identified using a swarm intelligence approach called Glowworm Swarm Optimization. These dominant features are then trained using a neural network with a regression function. Finally, the performance of the proposed method will be evaluated based on accuracy, sensitivity, and specificity and compared to existing approaches, such as SURF, with different variations for facial recognition with and without masks.

Keywords Facial recognition · Pandemic environment · With masks · Without masks · Cropping · Neural network

✉ Kosuri Naresh Babu
naresh.kosuri@gmail.com

¹ Dept of CSE-AIML, Geethanjali College of Engineering and Technology, JNTUK, Kakinada, Hyderabad 501301, India

² Dept of IT, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada 520007, India

Abbreviations

| | |
|--------|---|
| GRNN | Generalized Regression Neural Network |
| AHE | Adaptive Histogram Equalization |
| HMF | Histogram Modification Function |
| SIFT | Scale-Invariant Feature Transform |
| SURF | Speeded-Up Robust Features |
| U-SURF | Unsupervised Speeded-Up Robust Features |
| HGSO | Hybrid Glowworm Swarm Optimization |
| GSO | Glowworm Swarm Optimization |

1 Introduction

The human face is recognized as one of the most valuable characteristics in biometrics, and it has been widely employed in a various application, including security, military, and education. However, the majority of presently in use face recognition systems make assumptions about ideal imaging circumstances, such as the ability to acquire fully featured photos of sufficient quality for use in the recognition process [1–3]. As the number of new infections during the COVID-19 pandemic will be significantly impacted by the uncovered face.

Traditional non-occluded faces which include all specific facial areas, like those of the eyes, nose, and mouth, are portrayed by face recognition systems (in collaborative circumstances). The widespread mandate that individuals wear respirators face masks in social situations, prompted either by COVID-19 pandemic strategy, has developed a need to comprehend how cooperative face recognition technology accommodates obscured faces, including when only an periocular area is visible. In this research, we implement a periocular recognition system for unregulated wild situations to overcome this issue (i.e., masked faces) [4, 5].

When a human identifies a face, he or she considers not only the overall aesthetic pattern but also semantic details like gender, race, age, etc. to determine yet if the face refers to a certain recognised individual. Consequently, it makes sense to assume that semantic data is beneficial for the job of thermographic identification. This issue also arises in face authentication image processing systems. Numerous research have been done to identify faces in difficult situations because of variations in stance, light, picture quality, occlusion, etc. [6]. However, merely in the last two years has research on mask-based face identification exploded. Throughout the epidemic When facial recognition algorithms fail to distinguish faces as a result of the increased popularity of wearing a facial mask to stop the transmission of the coronavirus, the occlusion issue is clearly emphasised. Even the facial recognition on our mobile device is unable to recognise the owners. In border control systems, airports, and other locations with enhanced security, this issue is more acutely exposed. Occlusions occur often in natural settings and are particularly difficult and troublesome in many aspects of digital vision and object recognition since they may happen to any item in an unrestricted environment and they completely ruin the subject's features [7]. Due to the unexpected nature of occlusions, face identification under occlusions continues to be a significant difficulty. Due to the fact that the occluded region of a face picture might vary in location, size, and form, facial identification presents a highly difficult problem. In unrestricted contexts, facial occlusions are inevitable, and there are untold millions of possible occlusion scenarios. It is not practical to gather a large dataset containing every conceivable occlusion condition in order to develop a deep neural system. There are

several possibilities that may happen. Four different types of facial occlusions may be distinguished.

- o Face-covering items, such as hats, facemasks, eyeglasses, and hair.
- o Occlusions from the outside: covered with palms or other unrelated things.
- o Faces that were only partly recorded because of the camera's narrow field of vision.
- o Artificial occlusions: random salt-and-pepper noise, random white or black rectangles [8, 9].

The Deep Learning idea of Convolutional Neural Network (CNN) has been extensively used to pattern recognition. Researchers have conducted tests to categorise faces, whether they are hidden or not, using CNN. InceptionV3 and Exception are the two most notable models after a survey of five different CNN architectures is used in research [10–12]. Analysis revealed a strategy that was thought to be more effective by omitting the mask's coverage of the face. Additionally, they used VGG-16, Alex Net, and ResNet-50 pre-trained CNNs to extract features from the obtained regions (eye and forehead area). The "Real-World-Masked-Face-Dataset" testing results demonstrate the best accuracy by 91.3%. The MTCNN face identification model was suggested as a method for finding face masks in videos. The next step is for a MobileNetV2 object detector to determine if a face is covered or not. The accuracy of the face detection was 81.84%, while the accuracy of the face mask detection was 81.74%. We need to know the position of this face in order to solve the detection issue, which calls for a class that can decide whether face is hidden or not. Because previous observations have only been created for faces without masks, mask detection is difficult. Many conventional detection methods that are based on carefully built feature extraction are often utilised [13–15].

This paper offers a biometric system for humans, despite of whether they wear a face mask or not, in keeping with the works listed in the bibliography. The work has been divided into four parts for this purpose: Sect. 2 includes the materials and techniques, Sect. 3 displays the findings, and Sect. 4 gives the commentary.

2 Related work

In [16] research, researchers suggested a mask detection technique that relies on HOG (Histogram of Gradient) attributes classifier and SVM (Support Vector Machine) to establish whether the face is masked or not. The proposed method was tested using over 10,000 randomly selected images from the Masked Face-Net database, and it correctly classified 98.73% of the tested images. Furthermore, a novel geometrical features extraction approach that relies on Contourlet transform was presented in order to extract sufficient characteristics from partly occluded face pictures. When evaluated against the Masked Face-Net database's 4784 properly masked face photos, the approach has an identification accuracy of 97.86%. The Covid-19 epidemic led to the development of the online education system. The main goal of this article is to map the relationship between instructional strategies and online student learning. For formative evaluations to examine student comprehension in a real-world setting, face-to-face evaluation approaches are reasonably fast and simple. The majority of research show a connection between a person's emotions and facial expressions. Teachers often get daily input from the students to improve the teaching–learning process. The process may be made more participatory and teaching techniques can be improved with the help of this input. It's important to recognise and comprehend people's emotions while studying virtually. Facial recognition algorithms may be used to extract useful information from internet platforms. Exams are administered via a linked online course with students, and the findings indicate that this method works well [17]. During the

COVID-19 epidemic, going to wear a facial mask is customary and suggested in confined environments like workplaces.

A face mask is regarded as a partial occlusion in face recognition, which reduces recognition precision. The occlusion factor caused by various face mask designs is the main topic of this research. The purpose of this research is to lessen how much a facial recognition system will be impacted by face masks. The photos of the faces were covered with a fake face shield and black occlusions by the authors (FI). In order to extract face embeddings, System is set, a narrow deep neural network, was employed. With the help of a support vector machine, the faces were categorised. Authors tested several situations utilising various training and testing sets that include various mask designs. In a controlled setting, it demonstrated effectiveness in identifying obscured reduced FI with such estimated overall accuracy of 98.93% [18].

Recent rapid advancements in deep learning have produced some of the most encouraging outcomes for face recognition systems. However, when faced with difficulties like fluctuating lighting conditions, poor resolution, facial emotions, position variation, and occlusions, they perform at much below tolerable levels. A common belief is that facial occlusions are among the most difficult issues to solve. Particularly when the occlusion covers a significant portion of the face since it obliterates several official features [19]. The implementation of YOLOV3 as either a one-stage detector produced a bounding box for the face region and a class prediction at the same time. We examined the effectiveness of three pre-trained models for facial recognition: ResNet152V2, InceptionV3, and Exception. Results for the mask detection were encouraging, with MAP values of 0.8960 for training and 0.8957 for validation. For face recognition, we choose the Exception model since it has less parameters and equivalent quality to ResNet152V2. On face photos bigger than 100 pixels, exception obtained a low return loss inside the validating of 0.09157 with 100% accuracy. Overall, the system shows promise and is capable of recognising faces, even when hidden by masks [20]. The present research focuses on a measuring environment's initial idea and assessment. The created idea was selected to use a multidimensional approach. Electrocardiography (ECG) and electrodermal activity (EDA), as well as forearm and neck electromyography, were the psychophysiological measures that were investigated (EMG). The visual information was also captured on video. The findings of a preliminary assessment demonstrate that such combination of the several measures was effective and encouraging for use in further study [21]. The authors of [22] suggest a deep learning-based method for identifying online learners' real-time involvement using facial expressions. This is accomplished by analysing the students' facial expressions during the digital learning session to categorise their moods. The engagement index (EI), which predicts the engagement states "Engaged" and "Disengaged," is calculated using the information on face expression recognition. The improved statistical machine learning technique for actual engagement detection is determined by evaluating and comparing several deep learning models, including Inception-V3, VGG19, and ResNet-50. The overall effectiveness and accuracy of the suggested approach are evaluated using several benchmarked datasets as FER-2013, CK+, and RAF-DB. On benchmarked datasets as well as our customer dataset, experimental findings demonstrated that the suggested scheme reaches an accuracy of 89.11%, 90.14%, and 92.32% for Inception-V3, VGG19, and ResNet-50, respectively. In real-time learning circumstances, ResNet-50 surpasses the competition including a validity of 92.3% for classifying facial emotions. The aim of this systematic review was to develop an automated face detection and identification system for use in security and healthcare applications. A face recognition virtual environment was made utilising the Raspberry Pi device and the Open VINO toolbox, and applications like the Interactive Face Detection

Demo and the Human Pose Estimation Model were tested there. Initially, the idea was to gather photographs of initial random individuals for the Raspberry Pi to learn from, and then have the subjects appear in front of the gadget to check whether the linked camera could recognise them. However, there were a few subjects available because of the COVID-19 epidemic. Therefore, because there is a broad range of material available via Twitter, Instagram, and YouTube, pictures of the 13-member K-Pop group Seventeen were utilised. Additionally, members often alter their hair colour and style, so a range of photos might be utilised to thoroughly verify the device's accuracy. Each member was exposed to 6–8 distinct photos for the Raspberry Pi's camera to recognise in order to analyse it. The product's image database had 4–5 of these pictures, while just 2–3 were unknown. Among the 100 total images, the Raspberry Pi was able to identify faces in 77 (or 77%) of them, but in 23 (or 23%) of them. 24 (31.17%) of the 77 photos the Raspberry Pi determined to feature faces were properly identified, 52 (67.53%) were mistakenly identified, and 1 (1.3%) was distinguished as unknown. Despite the fact that the gadget was mainly incorrect, it was found that adjustments to the sensitivity with which the camera scans and recognises face characteristics would increase the device's accuracy. A way for incorporating the gadget into people's everyday life as a self-care notification system was developed during the experimental process, and the notion of using the technology to identify indications of emotional stress in teens was developed [23]. It becomes more and more challenging for facial recognition algorithms to recognise persons wearing masks as everyone now covers their faces to prevent the spread of the COVID-19 virus. Face-recognition algorithms created before the epidemic often fail in this situation, thus it is important to understand how they function when faced with obscured faces. The goal of this research is to create a novel, simple Convolutional Neural Network-based solution to address this problem. The suggested model provides accuracy that is equivalent to earlier, related models. Additionally, the suggested technique is employed to develop a reliable system to guarantee COVID-19 protocol adherence in a practical setting [24]. This study suggests a compact convolutional neural network (CNN) network that has been tuned for identification of involvement in a distance-learning setting using facial expressions. The Shuffle Net v2 architecture was chosen because it offers superior performance in contrast to other lightweight models and can readily adapt to mobile devices. The suggested model was developed, put to the test, assessed, and contrasted with existing CNN models. The outcomes of our study demonstrated that the highest performance for engagement recognition is delivered by an optimised model built upon this Shuffle Net v2 infrastructure with a modified activation function and the addition of an attention mechanism. Additionally, using the same database, our suggested model performs better in terms of engagement recognition than many research. Finally, this technique may be used to recognise student participation in mobile platforms for remote learning [25].

3 Proposed work

3.1 Problem statement

The confinement of individuals to their houses reduces output and slows down economic dynamism, two consequences of COVID-19 that are visible to the human eye. Even though, it should be mentioned that prioritising people's health before any constructive endeavour is vital in circumstances of health crisis, much like the one that is still being felt.

To prevent the expansion of this harmful virus, biosecurity precautions and social distance guidelines have been put in place. Additionally, the capacity of businesses, organisations, and other places has been constrained, emphasising as so telecommuting (in certain cases). As a result, organisations have put in place a variety of procedures, tactics, and techniques to safeguard people's integrity and health as they enter and remain in face-to-face work settings. As was already established, CNN has been a crucial piece of technology throughout this epidemic. The majority of methods have been used to diagnose the condition, although monitoring and prevention have also been discussed. Today, wearing a protective facial mask is a required precaution. As a result of covering the lips, nose, and cheeks, persons are now only identifiable by their eyes, brows, and hair. This presents a challenge for human eye, which is apt to notice similarities among multiple faces that have the same characteristics. Given how prevalent face recognition technologies are now, this issue also impacts computer systems. They are used to enter certain locations, access private apps, and unlock smartphones. Technology must adjust to these new circumstances since existing systems typically handle data from a person's complete face. All of this is done to keep the user's biosecurity while yet allowing them to go on with their tasks as comfortably as possible. There are methods that try to determine if individuals utilise it correctly, according to the literature. These efforts have produced excellent outcomes. However, no research has been done on exploiting biosecurity material for face recognition. All of this served as the impetus for the current inquiry, which presents a surveillance system with two techniques. The first step is to create a face classifier using a database of persons wearing and not wearing masks. The second talks about a face recognition algorithm that, in controlled conditions, enables automated identification of persons without taking off the face mask. This may be deployed as a low-cost access system for a building or a residence. Using open-source programming tools and basic features that cut down on computational costs ensures this. As a result, the prospect of making present facial recognition systems more adaptable to changing conditions has been established as a working hypothesis.

3.2 System model

The pipeline for face recognition begins with face detection, which is the initial stage in the process. There are certain restrictions in advanced unconstrained situations, although general face detection algorithms can be applied to some degree. A vast region being obscured makes it difficult to detect occluded faces in unrestricted contexts because pairwise similarity and intraclass variance increase. Utilizing adaptive technology, several strategies are created to effectively tackle the issue. It is suggested to create a system that can recognise a person's face whether or not they are wearing a mask. Two databases are required for the system to function effectively. The first database serves as classifier training and contains a large number of photos of persons wearing face masks and those who do not. The face recognition system is trained using the second, where individuals with and without biosafety material are present (face mask). The architecture employed is GRNN with such goal of having superior accuracy and resilience. The datasets are received via either an image or a video.

The facial recognition pipeline consists of 4 parts.

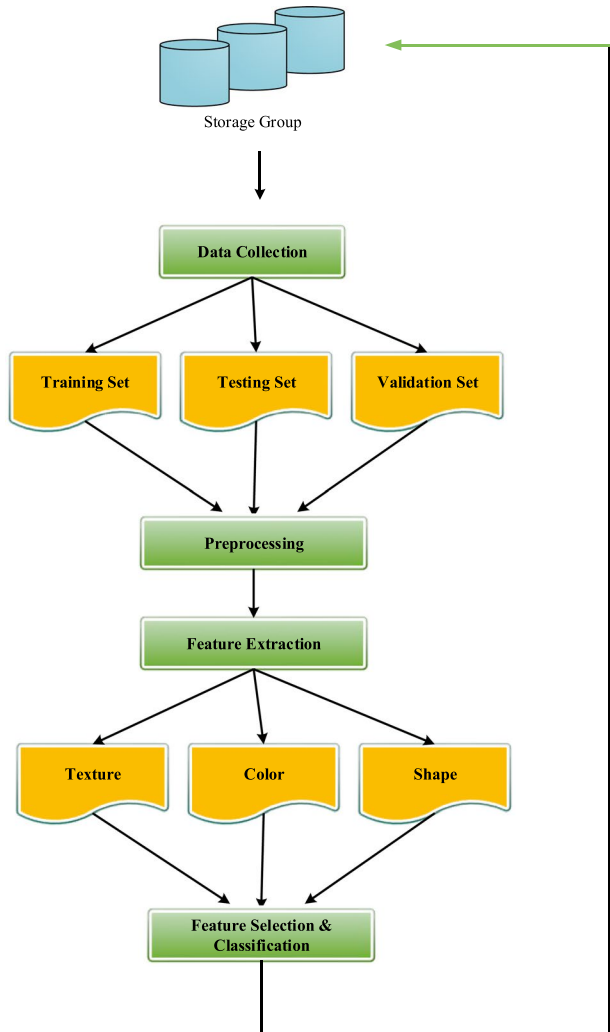
1. Identifying one or even more individuals in a picture is known as face detection.
2. Face Processing: Cropping, scaling, and alignments of the face.

3. Feature extraction is the process of removing key characteristics from a picture of a face.
4. Face Matching: Comparing extracted feature vectors to photos in the database.

Figure 1 illustrates the proposed work flowchart. Adaptive Histogram Equalization (AHE) methods offer a sophisticated approach to enhance the contrast of images, particularly in scenarios where traditional histogram equalization techniques may not be optimal. By adjusting the reflectivity of individual pixels in a picture to utilize the maximum possible number of bins, AHE effectively amplifies the contrast in the enhanced image.

One fundamental concept in histogram processing is the normalized histogram, which represents the proportion of pixels with potential intensities relative to all the pixels in an image. This normalized histogram provides valuable insights into the distribution of pixel values and is instrumental in the enhancement process. Normally the pixel values in medical pictures may be homogeneous, and the AHE enhances the noise in such areas. Contrast

Fig. 1 Proposed work flowchart



Limited Adaptive Histogram Equalization (CLAHE) may be used to alleviate the over amplification issue. By reducing the contrast of each pixel in the surrounding areas, the CLAHE develops a transformation function. The values of all the pixels combine to generate a slope in the Cumulative Distribution Function (CDF). The slope of the CDF is exactly proportional to the slope produced by the transformation function.

The pooling layer is a straightforward and efficient filtering method for removing impulsive noise from a picture. In order to encompass the pixels whose sample mean is to be determined; it employs a single layer of changeable size. To get the centre value, the values of the pixels within the window are sorted in ascending order. The discovered median value is used to swap out the image's centre pixel. The median filtering filter is employed when the amount of unwanted noise is larger than two. The adapted median filter helps identify impulsive noise, smooth out other disturbances, and reduce visual distortion. It identifies the noise that is present in a picture by comparing the values of every single pixel to those of its neighbours. The average pixel value is then used in lieu of each pixel's value.

In this study, the provided picture is pre-processed using unique filtering method called the Hybrid Median Filter (HMF). It is a logical progression from non-linear hybrid filters of the rational kind. a technique for creating a very condensed and accurate depiction of a picture that has a forefront and a background. The impulse noise in the photos is efficiently eliminated using HMF, a non-linear class windowed filter. The dual filter is shown to have superior edge maintenance qualities than the traditional median filter. In this method, the centre pixels' new value is determined by the median values of the nearby pixels, which are sorted by intensity. HMF is a modified median filter that eliminates noise more effectively than either the median filter. The primary goal of noise reduction is to reduce noise while maintaining picture information. Moreover, it alters the pixel values to create a new picture from an original.

The picture is subjected to many applications of the median approach while changing the window shape and considering the midpoint of the collected median values. The diagonal "D" pixels' median value "MD" and the azimuth and elevation "R" pixels' median value "MR" are determined. The middle pixel M and the average of the two predicted values are used to calculate the filtered value. The resulting data of the azimuth and elevation pixels is shown in Fig. 2. the resulting data of the orthogonal pixels is shown in Fig. 2(b) and the centre pixel is shown in Fig. 2(c). The following provides an explanation of this.

$$\begin{bmatrix} D & * & R & * & D \\ * & D & R & D & * \\ R & R & DMR & R & R \\ * & D & R & D & * \\ D & * & R & * & D \end{bmatrix}$$

3.3 Hybrid Median Filtering Technique

Step 1: Calculate the median values for the horizontal and vertical pixels 'M_R'

Step 2: Calculate the median values for the diagonal pixel 'M_D'

Step 3: Find middle pixel value.

Step 4: Calculate the filter value using the mean value of M_R, M_D and middle pixel value M

Step 5: Derive the filter value by calculating the median of M_R, M_D and M

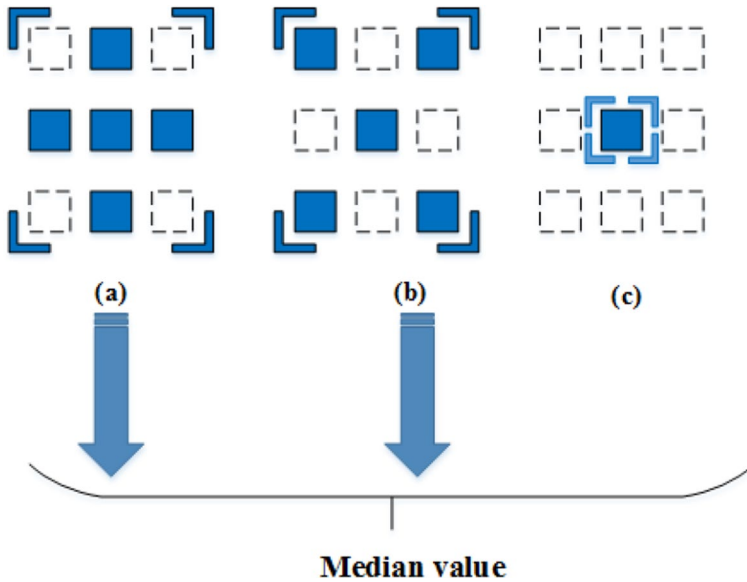


Fig. 2 (a). Median value of horizontal and vertical pixels (b). Median Value of diagonal pixels and (c). Middle pixel

$$F = \text{median}(MR, MD, M)$$

When compared to the other filtering techniques, the HMF provides the following advantages:

- No reduction in contrast across steps, since the output values available contains only of those present in the neighbourhood.
- It does not shift the boundaries, as can happen with the conventional smoothing filters.
- Since, the median is less sensitive than the mean values.

3.4 Feature extraction

The procedure of characterising an image's attributes or a group of its attributes that are effective and useful for representing information in the picture is referred to as attribute extraction. The categorization and examine of pictures may both benefit from the retrieved characteristics. The initial feature set, which comprises a starting set of measured characteristics, is where the pertinent features are generated from.

The learned characteristics are made easier to understand since the extracted features are instructive and non-redundant. The image's dimensionality reduction and feature extraction are connected. The elimination of duplicate data from a big feature set should be accomplished utilizing one of the segmentation techniques or strategies since processing such a collection is highly important. The phrase "feature vector" also refers to the condensed collection of features, which is given in Fig. 3.

When comparing to other descriptors, SIFT performs well. It appears to provide strong distinguishing power while avoiding the consequences of localization mistakes in terms of

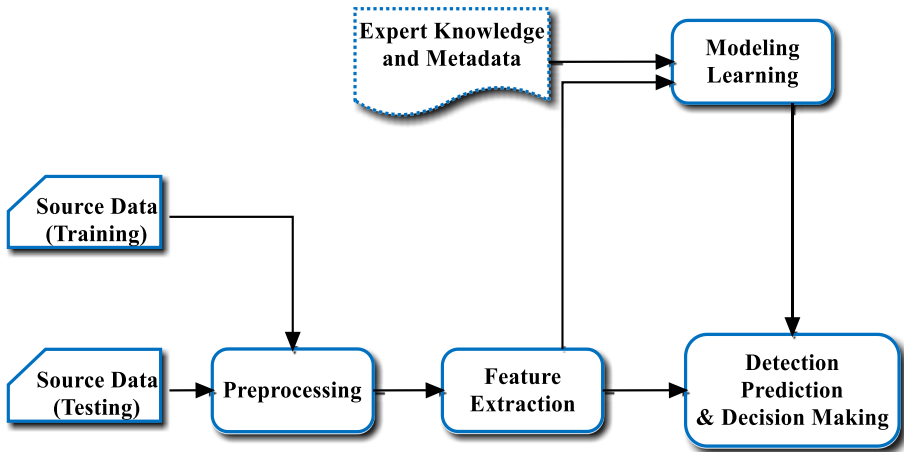


Fig. 3 Process of feature extraction

size or space due to the combination of coarsely located knowledge and the distributions of gradient-related characteristics. Photometric shifts are less noticeable when gradients are used with appropriate respective strengths and orientation. Similar features serve as the foundation for the proposed SURF descriptor, which has its complexity further reduced. Fixing a replicable orientation based on data from a circular area around the object of interest is the first stage. The SURF descriptor is then extracted from a square area that has been constructed and is aligned to the chosen orientation. Now, each of these two phases will be described. Additionally, we also suggest an upright variation of our description (U-SURF) that is quicker to calculate and better suited for situations where the camera stays mostly horizontal since it is not invariant to picture rotation.

The greatest value of the determinants of the Hessian matrix is used to discover the characteristic spots for the SURF, which are based on the Hessian approximation matrix. The picture is convolved with the second element Gaussian differential template for the computation of the Hessian matrix. The convolution of the templates and the images is changed into a box filter operation to simplify the second order Stochastic dispersion templates, resulting in a template that is only made up of a few rectangular areas.

By integrating the picture, the responsiveness value of the boxes filter may be easily solved. The following is a definition of the particular integral image:

$$I_{\Sigma}(x, y) = \sum_{i=0}^x \sum_{j=0}^y I(i, j) \quad (1)$$

By incorporating the image, it is simple to resolve the response of something like the box filtering, or the grey value to a rectangular region of any size. As a result, the following formula is used to get the fundamental amount of the approximated Hessian matrix:

$$\det(H_{\text{approx}}) = D_{xx}D_{yy} - (0.9D_{xy})^2 \quad (2)$$

where $\det(H_{\text{approx}})$ is the determinant value of approximate Hessian matrix. The responsive images of feature points identification at a particular scale is created by approximation the determinants value of something like the Hessian matrix to represent the value of the

dependent variable of features extracted at a specific point x in the image and exploring all pixels the image's points. The template size of various box filters is used to create the pyramid picture of the response to multi-scale feature points. Through the use of a pyramid picture, 3D non-maximum suppressing is performed out, and then sub-pixel extrapolation is used to determine the precise positions of the real extreme examples.

$$\hat{\chi} = -\frac{\partial^2 H^{-1}}{\partial x^2} \frac{\partial H}{\partial x} \quad (3)$$

For each sub-block, Haar template of size $2s$ is used to calculate the response value, and the response value is counted, and the 4-dimensional vector $(\sum dx, \sum |dx|, \sum dy, \sum |dy|)$, dx and dy are the response value of the Haar wavelet in the x and y directions. Therefore, SURF feature points are described as $4 * 4 * 4 = 64$ -dimensional vectors. Finally, the response was summed and normalized.

In order to provide different feature description, SURF leverages integral pictures. It establishes the main direction before features extraction. The wavelet size is set to $4 * s$ and the sampled step is s , where s is the scale of the facial landmarks to be identified. The Wavelet coefficients response value in the image's x and y planes are computed in the neighbourhood with a radius of $6 * s$. Be using the Stochastic weighting scheme of $=2 * s$ to give the wavelet response value some weight. A fan-shaped slider window with a size of $/3$ that is centred on the feature point is rotated to move the windows at a step length of 0.2 radians in order to acquire the primary direction value. The response values of the images Haar wavelet inside the moving window are alternately summed in the horizontal and vertical directions. The primary orientation corresponds to the orientation of the longest set of vectors created by the two replies added together. The $20 * s * 20 * s$ image was segmented into $4 * 4$ sub blocks and along main direction, and the response values for every subblock were calculated using the Haar templates with a size of $2 * s$.

3.5 Feature selection

In image acquisition, machine learning, and data mining applications, feature extraction is the process of lowering the dimensionality of the retrieved feature set. In the feature selection step, the feature set derived out from feature extraction procedure is further condensed. From a basic feature set that includes a huge number of features, the feature selection selects the best features. In order to enhance the image's quality, it is employed to eliminate noisy characteristics. It is also known as the attribute selection process or differential subset selection process since the needed attributes are extracted from the initial feature set. The primary distinction between feature extraction and feature selection is the creation of new features; in contrast, feature selection simply selects existing features.

3.6 Classification

The crucial step in digital image analysis is picture categorization. To divide all of the adjacent pixels in a digital picture into different groups or classes is the primary goal of categorization. It refers to the process of dividing the photos into those whose pixel values are known and those whose identities are unknown. In image classification, non-overlapping

portions of the decreased feature space are separated into distinct classes and given distinctive labels. The reliability of the categorized picture is compared to the uncategorized image after the classification of the photos.

3.7 Updating the result in excel

To perform this operation, a database will be created with the person names in columns and date as rows. Once the image was identified from the classification process, the attendance will be marked as present in the excel sheet or database. This system will be helpful for performing an automated attendance system and it saves time.

The collection of facial expression characteristics and the choice of classifiers are more difficult with conventional face recognition, as well as the detection performance is not very high. Together with fully convolutional network's further development from handwritten digit identification to facial recognition, a facial recognition system to evaluate CNN which is given in Fig. 4. Two factors dominate the technique. One method is to watch how the network is affected by altering the number of neurons within the hidden layer, and the other is to watch how the network is affected by altering the quantity of feature mappings in the convolution layers 1 and layer 2. The most effective CNN model has undergone several experimental testing's with various sizes. The model has the ability to automatically

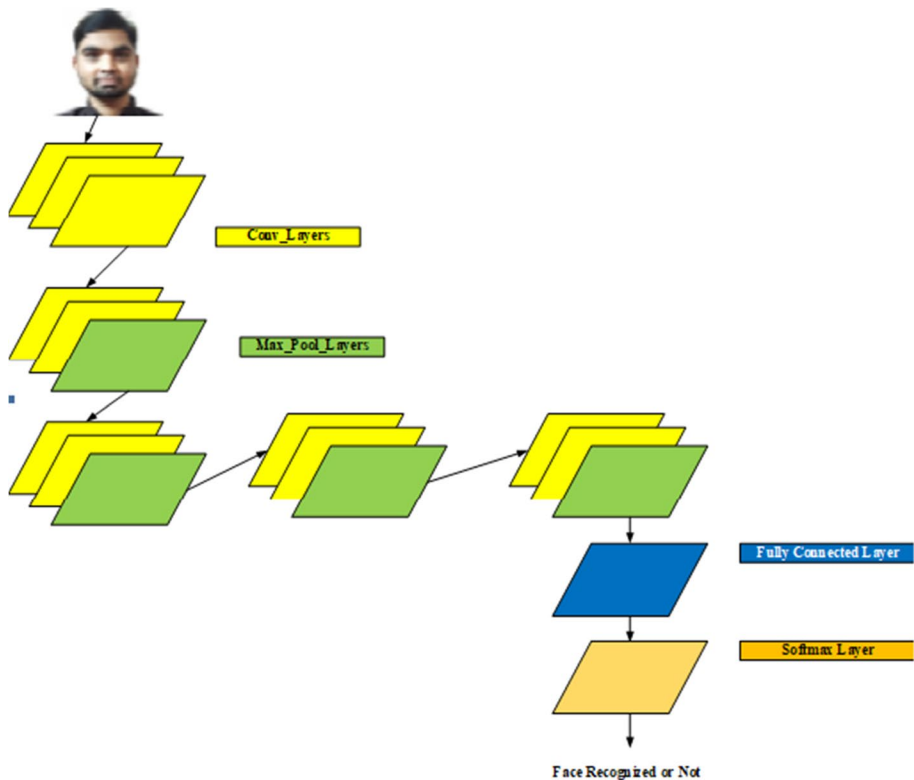


Fig. 4 CNN for Face Recognition

extract and identify characteristics from face images. For face recognition, the proposed method and softmax classification can speed up train converge and do more. Effectively increase accuracy while avoiding overfitting by using the Dropout technique.

4 Findings and discussion

The tests are conducted using a recollection metric, Finesse, F1-score, and the accompanying macro average and weighted average, with the goal of illustrating the possible utility of the datasets. Utilizing these measures has as its goal evaluating the system from many angles. Recall and precision show how well the model can identify genuine positives. Recall takes into account both the model's accuracy in identifying false positives and false negatives. False positives happen when an item is identified as a face in this example of face detection using masks.

4.1 Dataset

Different people images without mask at different illumination and angles were captured for feature extraction and training, which is given in Fig. 5.

Figure 6 illustrates the pre-processing result. The above process was common for both training and test set with and without masks.

In Fig. 7 row represents the image, column represents the features and feature selection is implemented using HGSO algorithm Figs. 8, 9, 10, 11, 12, 13 and 14.



Fig. 5 Sample images from dataset

Fig. 6 Preprocessing Results (a). Input image, (b). Gray conversion, and (c). Median filtering



Variables - X

X

150x12 single

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|----|--------|--------|---------|--------|--------|--------|--------|--------|--------|--------|--------|---------|----|
| 1 | 5.1911 | 1.3708 | -0.9600 | 3.3481 | 0.2994 | 7.1345 | 0.0977 | 0.9739 | 0.9531 | 0.2026 | 1.9418 | 38.1333 | |
| 2 | 4.6349 | 1.1542 | -0.8864 | 3.3205 | 0.3162 | 6.9932 | 0.0737 | 0.9750 | 0.9641 | 0.2492 | 1.8741 | 35.2000 | |
| 3 | 5.6526 | 2.8026 | -1.1134 | 3.3004 | 0.2450 | 7.4485 | 0.1148 | 0.9840 | 0.9461 | 0.1731 | 1.9985 | 31.0667 | |
| 4 | 5.2813 | 4.7656 | -0.7377 | 2.1847 | 0.1889 | 7.6276 | 0.1718 | 0.9857 | 0.9264 | 0.1377 | 2.0401 | 29.3333 | |
| 5 | 4.3612 | 2.4544 | -0.9177 | 2.7261 | 0.2338 | 7.0498 | 0.0704 | 0.9874 | 0.9653 | 0.2060 | 2.0144 | 20.4000 | |
| 6 | 5.2017 | 4.8678 | -0.6723 | 2.0650 | 0.1837 | 7.6502 | 0.1687 | 0.9863 | 0.9261 | 0.1337 | 2.0371 | 28.9333 | |
| 7 | 5.8009 | 1.6713 | -1.3609 | 4.5238 | 0.2844 | 6.8987 | 0.0868 | 0.9785 | 0.9595 | 0.2899 | 2.0873 | 35.7333 | |
| 8 | 4.0207 | 3.0833 | -0.4442 | 1.8490 | 0.1863 | 7.4562 | 0.1016 | 0.9858 | 0.9501 | 0.1616 | 1.8845 | 35.3333 | |
| 9 | 5.7287 | 1.9326 | -1.3744 | 4.4183 | 0.2749 | 6.9642 | 0.0845 | 0.9819 | 0.9606 | 0.2796 | 2.0917 | 30.5333 | |
| 10 | 4.7424 | 4.3592 | -0.5837 | 1.9235 | 0.1772 | 7.5910 | 0.1312 | 0.9874 | 0.9350 | 0.1483 | 1.8273 | 32.5333 | |
| 11 | 4.6941 | 1.6027 | -1.0221 | 3.5106 | 0.2686 | 7.1779 | 0.1301 | 0.9670 | 0.9368 | 0.2056 | 1.8673 | 33.8667 | |
| 12 | 4.7045 | 3.4959 | -0.8011 | 2.3834 | 0.2085 | 7.4744 | 0.1041 | 0.9872 | 0.9495 | 0.1677 | 1.9398 | 30.2667 | |
| 13 | 4.4708 | 2.8786 | -0.4659 | 2.2031 | 0.1778 | 7.6127 | 0.1075 | 0.9852 | 0.9467 | 0.1258 | 1.8040 | 34.6667 | |
| 14 | 4.8385 | 4.3990 | -0.6741 | 2.1061 | 0.1814 | 7.5310 | 0.1328 | 0.9872 | 0.9349 | 0.1385 | 1.8912 | 30.6667 | |
| 15 | 4.6904 | 2.5562 | -0.3790 | 1.9355 | 0.1937 | 7.5307 | 0.0960 | 0.9855 | 0.9522 | 0.1424 | 1.8484 | 50.4000 | |
| 16 | 4.5216 | 3.3700 | -0.2715 | 2.0414 | 0.1530 | 7.7993 | 0.1367 | 0.9935 | 0.9353 | 0.1102 | 1.8291 | 48.4000 | |
| 17 | 5.0783 | 4.0367 | -0.8914 | 2.4892 | 0.2051 | 7.3913 | 0.1250 | 0.9865 | 0.9393 | 0.1851 | 1.9454 | 33.8667 | |
| 18 | 3.7292 | 1.0233 | 0.6742 | 5.5624 | 0.3747 | 6.8142 | 0.0905 | 0.9641 | 0.9580 | 0.2844 | 2.0134 | 34.8000 | |
| 19 | 4.3567 | 2.3234 | -0.3032 | 2.2591 | 0.1973 | 7.5019 | 0.1073 | 0.9821 | 0.9498 | 0.1521 | 1.9021 | 40.5333 | |
| 20 | 4.6909 | 4.1204 | -0.2516 | 1.4903 | 0.1953 | 7.5622 | 0.1211 | 0.9891 | 0.9416 | 0.1325 | 1.9448 | 44.5333 | |

Fig. 7 Feature Extraction Results

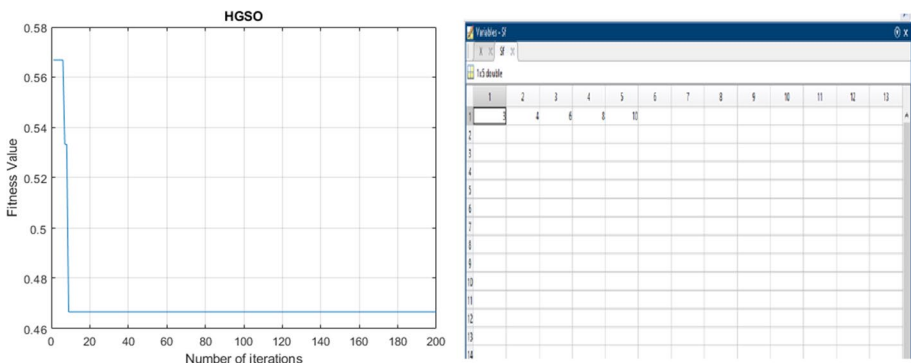


Fig. 8 Final Optimum Feature Set

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|----|---------|--------|--------|--------|--------|---|---|---|---|----|----|----|----|
| 1 | -0.9603 | 3.3481 | 7.1345 | 0.9739 | 0.2026 | | | | | | | | |
| 2 | -0.8864 | 3.3205 | 6.9932 | 0.9750 | 0.2492 | | | | | | | | |
| 3 | -1.1134 | 3.3004 | 7.4485 | 0.9840 | 0.1731 | | | | | | | | |
| 4 | -0.7377 | 2.1847 | 7.6276 | 0.9857 | 0.1377 | | | | | | | | |
| 5 | -0.9177 | 2.7261 | 7.0498 | 0.9874 | 0.2060 | | | | | | | | |
| 6 | -0.6723 | 2.0650 | 7.6502 | 0.9963 | 0.1337 | | | | | | | | |
| 7 | -1.3609 | 4.5238 | 6.8987 | 0.9785 | 0.2899 | | | | | | | | |
| 8 | -0.4442 | 1.8490 | 7.4562 | 0.9858 | 0.1616 | | | | | | | | |
| 9 | -1.3744 | 4.4183 | 6.9642 | 0.9819 | 0.2796 | | | | | | | | |
| 10 | -0.5837 | 1.9235 | 7.5910 | 0.9874 | 0.1483 | | | | | | | | |
| 11 | -1.0221 | 3.5106 | 7.1779 | 0.9670 | 0.2056 | | | | | | | | |
| 12 | -0.8011 | 2.3834 | 7.4744 | 0.9872 | 0.1677 | | | | | | | | |
| 13 | -0.4659 | 2.2031 | 7.6127 | 0.9852 | 0.1258 | | | | | | | | |
| 14 | -0.6741 | 2.1061 | 7.5310 | 0.9872 | 0.1385 | | | | | | | | |
| 15 | -0.3790 | 1.9355 | 7.5307 | 0.9855 | 0.1424 | | | | | | | | |
| 16 | -0.2715 | 2.0414 | 7.7993 | 0.9835 | 0.1102 | | | | | | | | |
| 17 | -0.8914 | 2.4892 | 7.3913 | 0.9965 | 0.1851 | | | | | | | | |
| 18 | 0.6742 | 5.5624 | 6.8142 | 0.9641 | 0.2844 | | | | | | | | |
| 19 | -0.3032 | 2.2591 | 7.5019 | 0.9821 | 0.1521 | | | | | | | | |
| 20 | -0.2516 | 1.4903 | 7.5622 | 0.9891 | 0.1325 | | | | | | | | |

Fig. 9 GRNN Training and Overall Evaluation of dataset

Fig. 10 (1). Input Image, (2). Gray Conversion, and (3). Median Filtering



Fig. 11 SURF Features



Fig. 12 Feature extraction and class

```

Command Window

\mOverall face_features
4.0564 4.4473 -0.0487 1.7856 0.1460 7.4173 0.1160 0.9889 0.9467 0.1264 1.8591 39.0647

\mOptimal face_features
-0.0487 1.7856 7.4173 0.9889 0.1264

\mThe face belongs to class
2

f: >>

```

Fig. 13 (1) Input Image, (2). Gray Conversion, and (3). Median Filtering



Fig. 14 SURF Features



- (i). Results for Without Mask.
- (ii). With masks.

From Fig. 15, the features for the test images will be extracted and the classes (that is the name of the person) will be determined. Then, the corresponding person attendance will be updated in the database as shown in the Table 1.

Fig. 15 Feature extraction and class

```

\Overall face_features
2.8610 1.8647 1.5916 4.4010 0.4110 6.6577 0.1713 0.9649 0.9491 0.3250 2.2953 33.0667

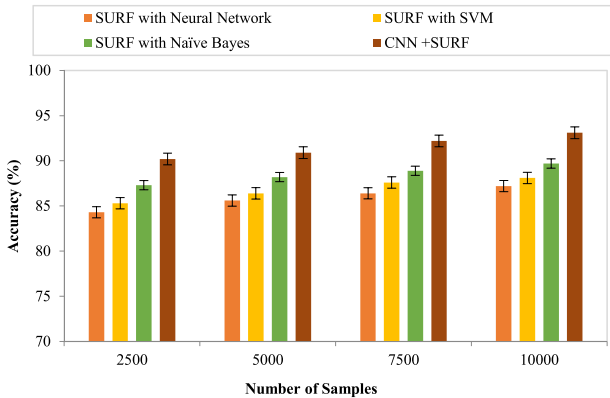
\Optimal face_features
1.5516 4.4010 6.6577 0.9649 0.3250

\The face belongs to class
2
    
```

Table 1 Person attendance

| S.NO | Month | Date | Aidai | anhu | aixin | baibahe | baijin |
|------|--------|------|---------|---------|---------|---------|---------|
| 1 | Dec-22 | 29 | present | present | present | absent | present |
| 2 | Dec-22 | 30 | absent | present | absent | present | present |
| 3 | Dec-22 | 31 | present | present | present | present | present |

Fig. 16 Accuracy vs. No of Samples



Because of the most important features being extracted from the final convolution layer of the pre-trained models, as well as the high efficacy of the presented GSO paradigm, which provides a lighter weight and much more discriminatory power compared to classical CNN with SoftMax Activation Function, the obtained high accuracy particularly in comparison to other face recognizers, is made possible. The suggested method's great degree of generality also enables it to be used in application scenarios as it only considers the unmasked areas. Conversely, some approaches use generating network to reveal the masked face. Implementations should not use this method since it is a greedy task. One of them used the identical pre-trained algorithms with a different strategy while competing against all other cutting-edge recognizers. Using the identical pre-trained models, the suggested technique performed better than CNN using SURF-based method. This effectiveness might be attributed to the fact that pre-trained model' fc levels include more dataset-specific attributes since they are often trained on real-time datasets, which is a completely distinct dataset. As a result, this approach is not always appropriate for our purpose. Additionally, the suggested approach performed better in terms of training phase than earlier

Fig. 17 Precision vs. No of Samples

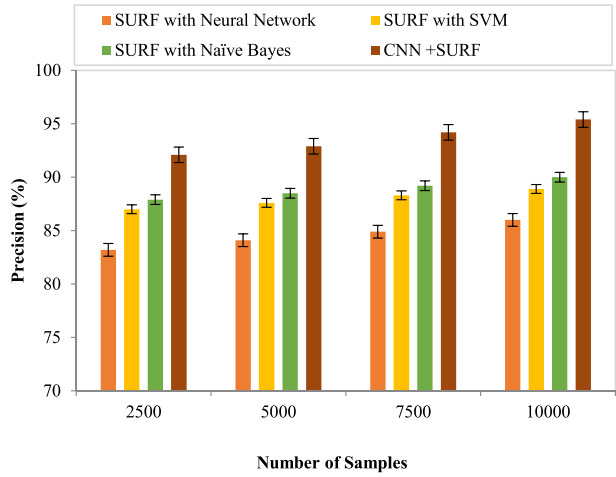


Fig. 18 Recall vs. No of Samples

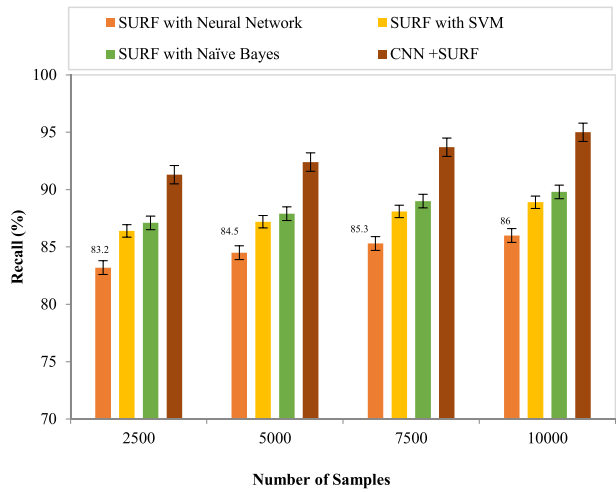
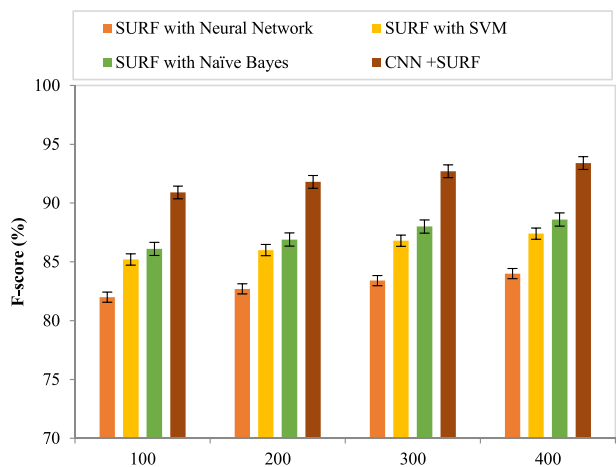


Fig. 19 F-score vs. No of Samples



approaches. The observed results provide substantial evidence that the CNN with SURF paradigms is a thin representation that strengthens the deep feature set's strong discriminatory power as a classification input, which is given in Figs. 16, 17, 18 and 19.

5 Conclusions and Future Scope

In this study addresses the pressing need for improved facial recognition algorithms capable of accurately identifying individuals both with and without masks. While deep learning and machine learning algorithms have shown success in facial recognition tasks, their performance tends to degrade significantly when faced with masked faces, primarily due to the occlusion of facial characteristics by the masks. In response to this challenge, our research proposes a novel facial recognition technique specifically designed to address the limitations posed by mask-wearing. By leveraging deep learning techniques and swarm intelligence, we developed a robust algorithm capable of identifying individuals even when wearing masks.

The proposed approach involves cropping images to isolate the common facial regions in both masked and unmasked faces. Features are then extracted using histogram properties, SURF, and SIFT features, capturing essential facial characteristics. The dominant features are determined using Glowworm Swarm Optimization (GSO), a swarm intelligence method, enhancing the recognition accuracy.

Furthermore, a neural network with a regression function is trained using these prominent features, enabling accurate identification of individuals regardless of mask presence. Performance evaluation using metrics such as accuracy, sensitivity, and specificity demonstrates the effectiveness of the proposed technique in real-world scenarios. Importantly, our research contributes to the growing body of literature on facial recognition technology, particularly in the context of post-COVID-19 challenges. By focusing on mask-aware facial recognition systems, we provide valuable insights and methodologies to address emerging needs in security, surveillance, and public health.

Looking ahead, future research efforts could explore the optimization of the proposed algorithm for real-time applications and large-scale deployment. Additionally, investigating the adaptability of the technique to different mask types and facial variations would further enhance its practical utility.

The performance of the suggested approach will then be assessed using accuracy, sensitivity, and specificity. For facial recognition software with and without masks, the suggested technique's progress will be compared to that of the currently used method, such as SURF with different variations.

Acknowledgements Not Applicable.

Funding Statement Author declared that no funding was received for this Research and Publication.

Data Availability The data used to support the findings of created new data set, this study is available from the corresponding author upon request.

Declarations

Ethical Approval This article does not contain any studies with human participant and Animals performed by author.

Conflicts of Interest The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

1. Hemathilaka, Susith & Aponso, Achala. (2021). An Analysis of Face Recognition under Face Mask Occlusions. 41–48. <https://doi.org/10.5121/csit.2021.111804>.
2. Aware, Mrunal & Labade, Prasad & Tambe, Manish & Jagtap, Aniket & Beldar, Chinmay. (2021). Attendance Management System using Face-Recognition. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*. 336–341. <https://doi.org/10.32628/CSEIT217370>.
3. Alexiou, Michail & Papadakis Ktistakis, Iosif & Goodman, Garrett. (2021). Towards a Masked Face Recognition Algorithm: A Novel Rule Based Hybrid Algorithm. 1–6. <https://doi.org/10.1109/SEEDA-CECNSM53056.2021.9566244>.
4. Baskaran, Kaaviya & P, Baskaran & V, Rajaram & Kumaratharan, Narayanaswamy. (2020). IoT Based COVID Preventive System for Work Environment. 65–71. <https://doi.org/10.1109/I-SMAC49090.2020.9243471>.
5. Arora M, Naithani S, Areeckal A (2022) A web-based application for face detection in real-time images and videos. *J Phys: Conf Ser* 2161:012071. <https://doi.org/10.1088/1742-6596/2161/1/012071>
6. Jiang, Andy & Sun, Yu. (2023). A Mobile Application to Mark Attendance using a Combined Back-end of the Firestore Database and Amazon AWS Services. 87–98. <https://doi.org/10.5121/csit.2023.130108>.
7. Pichetjamroen, Sasakorn & Rattanalerdnusrorn, Ekkachan & Vorakulpipat, Chalee & Pichetjamroen, Achara. (2021). Multi-Factor based Face Validation Attendance System with Contactless Design in Training Event. 637–640. <https://doi.org/10.1109/ECTI-CON51831.2021.9454779>.
8. Asaju, Christine & Vadapalli, Hima. (2021). Affects Analysis: A Temporal Approach to Estimate Students' Learning. 1–7. <https://doi.org/10.1109/IMITEC52926.2021.9714657>.
9. Mundial, I.Q., ul Hassan, M., Tiwana, M.I., Qureshi, W.S., & Alanazi, E.A. (2020). Towards Facial Recognition Problem in COVID-19 Pandemic. 2020 4rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM), 210–214.
10. Nursetiawati S, Josua DP, Atmanto D, Oktaviani F, Fardani AL (2020) Science Education in the Family Environment with the Experimental Method of Facial Cosmetics Plant Fertilization in the Covid-19 Pandemic Era. *Jurnal Pendidikan IPA Indonesia* 9:561–573
11. Tseng, F., Cheng, Y., Wang, Y., & Suen, H. (2022). Real-time Facial Expression Recognition via Dense & Squeeze-and-Excitation Blocks. *Human-centric Computing and Information Sciences*, 12. <https://doi.org/10.22967/HGIS.2022.12.039>
12. Awan, S.A., Ali, S., Hussain, I., Hassan, B., & Ashfaq Ashraf, S.M. (2021). Proficient Masked Face Recognition Method Using Deep Learning Convolution Neural Network in Covid-19 Pandemic. *International Journal of Circuits, Systems and Signal Processing*.
13. Savchenko A, Savchenko LV, Makarov I (2022) Classifying Emotions and Engagement in Online Learning Based on a Single Facial Expression Recognition Neural Network. *IEEE Trans Affect Comput* 13:2132–2143
14. Bradford, B., Yesberg, J.A., Jackson, J., & Dawson, P.A. (2020). Live Facial Recognition: Trust and Legitimacy as Predictors of Public Support for Police Use of New Technology. *The British Journal of Criminology*.
15. Srinivas, M.A., Kambam, V.R., Pinnu, S.K., Edla, S., & Komera, A.K. (2022). Facial Expression Recognition based Restaurant Scoring System Using Deep Learning. *Emperor Journal of Applied Scientific Research*.
16. Rifaae, M.A., Al Rawajbeh, M., AlOkosh, B., & AbdelFattah, F. (2022). A New approach to Recognize Human Face Under Unconstrained Environment. *International Journal of Advances in Soft Computing and its Applications*.
17. Enadula, S.M., SriramEnadula, A., & Burri, R.D. (2021). Recognition of Student Emotions in an Online Education System. 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT), 1–4.
18. Recto, I.J., & Devaraj, M. (2022). Synthetic Occluded Masked Face Recognition using Convolutional Neural Networks. 2022 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), 124–129.
19. Sekhar JNC, Domathoti B, Santibanez Gonzalez EDR (2023) Prediction of Battery Remaining Useful Life Using Machine Learning Algorithms. *Sustainability* 15(21):15283. <https://doi.org/10.3390/su152115283>
20. Nurhopipah, A., Rifai Azziz, I., & Suhuman, J. (2022). Behind the Mask: Detection and Recognition Based-on Deep Learning. (*Indonesian Journal of Computing and Cybernetics Systems*). Vol.16, No.1, pp. 67–78. <https://doi.org/10.22146/ijccs.72075>

21. Birkle, J., Weber, R., Möller, K., & Wagner-Hartl *, V. (2022). Psychophysiological Parameters for Emotion recognition – Conception and first evaluation of a measurement environment. *Intelligent Human Systems Integration (IHSI 2022)*. <https://doi.org/10.54941/ahfe100958>
22. Gupta, S., Kumar, P., & Tekchandani, R. (2022). Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models. *Multimedia Tools and Applications*, 1 - 30. <https://doi.org/10.1007/s11042-022-13558-9>
23. Bhattacharya, Pronaya & Handa, Anand & Zuhair, Mohd. (2021). Face Detection System for Health Care Units Using Raspberry PI. <https://doi.org/10.1109/ICIEM51511.2021.9445367>.
24. Mittal, Parangat & Pandey, Kartikeya & Tawani, Peeyush & Rohilla, Rajesh. (2021). CNN-based Person Recognition System for Masked Faces in a post-pandemic world. 1–6. <https://doi.org/10.1109/INCET51464.2021.9456416>.
25. Hu Y, Jiang Z, Zhu K (2022) An Optimized CNN Model for Engagement Recognition in an E-Learning Environment. *Appl Sci* 12(16):8007. <https://doi.org/10.3390/app12168007>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.