THE ASIAN BULLETIN OF BIG DATA MANAGMENT
Vol. 5. Issue 3 (2025)
https://doi.org/10.62019/abbdm.v5i3.365

ASIAN BULLETIN OF BIG DATA MANAGEMENT
ISSN (Print): 2959-0795
http://abbdm.com/
ISSN (online): 2959-0809

Optimized Deep Convolutional Neural Network for Robust Occluded Facial Expression Recognition

Muhammad Nauman, Muhammad Usman Javeed, Muhammad Talha Jahangir*, Shiza Aslam, Muhammad Khadim Hussain, Zeeshan Raza, Shafqat Maria Aslam,

Chronicle

Article history
Received: May 2, 2025
Received in the revised format: May 27, 2025
Accepted: June 20, 2025
Available online: July 30, 2025

Muhammad Nauman, Muhammad Usman Javeed, Shiza Aslam, Muhammad Khadim Hussain & Zeeshan Raza are currently affiliated with the Department of Computer Science, COMSATS University of Islamabad, Sahiwal, Pakistan.
Email: mr.nauman.edu@gmail.com
Email: usmanjavveed@gmail.com
Email: shizaaslam84@gmail.com
Email: khussain4912@gmail.com
Email: zeeshan.raza@yahoo.com

Muhammad Talha Jahangir is currently affiliated with the Department of Computer Science, MNS University of Engineering and Technology, Multan, Pakistan.
Email: mtalhajahangir@mnsuet.edu.pk

Shafqat Maria Aslam is currently affiliated with the School of Computer Science, Shaanxi Normal University, Xi'an, Shaanxi, China.
Email: shafqatmaria34@gmail.com

Corresponding Author*

Abstract

Occluded facial expression recognition (OFER) poses a formidable challenge in real-world applications, particularly in human-computer interaction and affective computing. Despite recent advancements, existing methodologies often struggle to maintain optimal accuracy under occlusion constraints. This study proposes a novel hybrid framework that synergizes handcrafted and deep learning-based features to enhance robustness and precision in emotion recognition. Specifically, we integrate Histogram of Oriented Gradients (HoG), facial landmark descriptors, and sliding window-based HoG representations with deep convolutional neural network (CNN) features, leveraging their complementary strengths. Our experimental design explores multiple feature fusion strategies, including CNN-based automated classification and a hybrid model incorporating Dlib-extracted landmarks with HoG-CNN integration. Comparative analysis against state-of-the-art approaches demonstrates that our multi-feature fusion technique significantly improves recognition accuracy, achieving a remarkable 96% accuracy on benchmark datasets such as RAF-DB and AffectNet. However, we observe a marginal decline in performance with increased dataset complexity, emphasizing the need for scalable solutions. This research underscores the efficacy of integrating handcrafted and deep learning-driven features, offering a promising direction for advancing occlusion-robust facial expression recognition in dynamic environments.

Keywords: Histogram of Gradients, Facial Expression Recognition, Occluded Faces, Emotion Detection, CNN.

INTRODUCTION

One of the newest breakthroughs in the forensics and computer vision fields is occluded facial recognition. In routine interactions, emotional expressions on the face play a significant role in the transmission and enhancement of information that voice cannot convey. Since it has numerous uses in fields including health protection, robots with emotional intelligence, monitoring driving fatigue, engaging game design, and machine-based analysis of facial expressions has becoming more popular in social marketing more interest in past few years [1]. Facial expression identification from an input image is one of the most difficult problems to solve, however by employing a We have enhanced the performance of facial expression identification using a Convolutional Neural Network based recognizer. Occluded facial expression identification is still a difficult Endeavor under several impossible circumstances, especially when faces are obscured. A face that is obscured by extraneous things makes it difficult to identify the face, such as a face covered by a scarf, one that is wearing glasses, a beard, a cap, or a mask. Other problems include lighting, posture, expressions, etc. Sunglasses, a hat, hands, hair, and other objects may very well obstruct some areas of the face Occluded Facial expression is used for nor verbal communication to understand others intentions. It has a lot of application like AI Tutor systems for feedback providing, Psychological Studies for pain detection, Human behaviour interpretation using robots, Mobile applications [2] for emotion insertion, automated security etc.

Issues has been risen by the researches in facial expression estimation like pose, illumination, low pixel density, baseline identification etc. Occlusion can significantly alter the optical appearance of the face and significantly impair FER system effectiveness [3]. Due to incorrect feature positioning, incorrect face alignment, or incorrect face registration, the presence of occlusion makes it more difficult to extract distinguishing features from obscured facial features. Previously conventional approaches used for facial emotion estimation Like SVM and automated approaches used for FEE like CNN [4].

Recent researches have shown that assembling of automated features and handcrafted give better results for a particular problem. So using this idea decided to merge different feature vectors (extracted by different techniques) to achieve better results. The goal is to build best possible network for Occluded Facial Emotion Estimation. This document will discuss about occlusion of handcrafted and automated feature to estimate facial emotion estimation. Experiment is carried out by extracting handcrafted features using Histogram of Gradients (HoG), Facial Landmarks, and automated features using CNN and FER2013dataset is used [5]. Later CNN was used to classify different images.

Facial Expression Recognition (FER) systems face significant challenges in accurately identifying emotions when confronted with occluded or obstructed facial regions. The blocking of different parts of the face by external objects such as sunglasses, hats, and masks or even lighting conditions change, varying poses, and density of the pixels interfere with the extraction of characteristic features that helps in establishing the emotion of the image. Contemporary systems like Support Vector machine (SVM) and single CNN based systems fail to succeed in beating these problems [6]. The inability of FER systems [7] to respond to the real-life conditions of occluded faces is the limitation of the absence of linkage between handcrafted and automated feature extraction methods.

The study seeks to establish an effective model of Occluded Facial Expression Recognition (FER) through the combination of automated and handcrafted features extraction methods. The specified method uses such features as Histogram of Oriented Gradients (HoG), facial landmarks, and deep features identified using Convolutional Neural Networks (CNNs) in order to increase emotion recognition accuracy during the occlusions caused by masks, scarves, sunglasses, and other obscures. This research aims at improving the FER performance by considering the changes in illumination, pose, and occlusion, hence leading to a more efficient estimation of emotions in the real world.

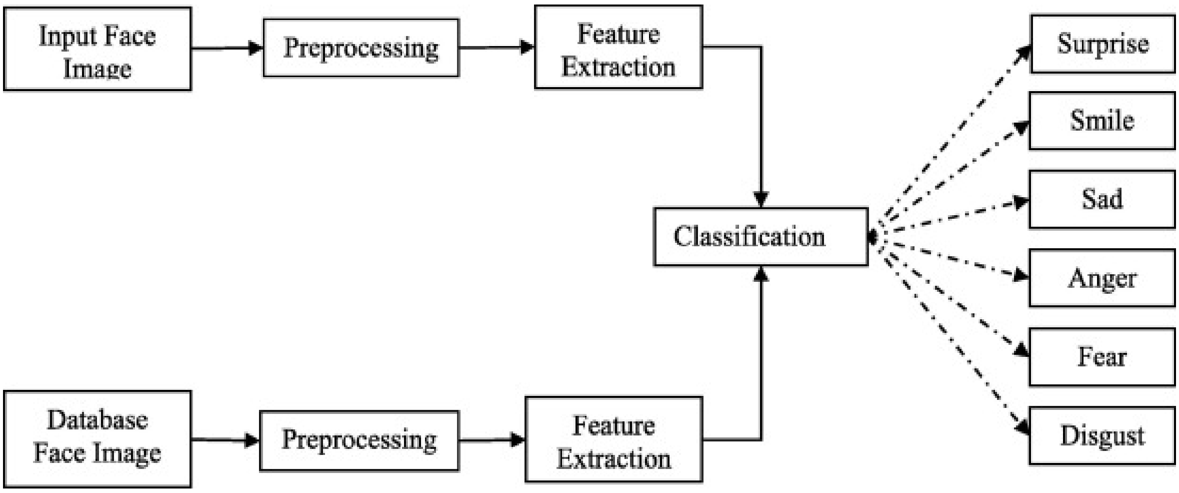


Figure 1.
Architecture of Facial Expression Recognition System

The research proposes a new framework that uses both handcrafted features, like Histogram of Gradients and Facial Landmarks, and automated features created with Convolutional Neural Networks (CNNs) in an attempt to enhance occluded Facial Expression Recognition (FER). By conducting the comparative analysis of various feature extraction methods and combining them, it is possible to identify the benefits of handcrafted and automated features integrations. Another study part is development and evaluation of a hybrid FER system with the FER 2013 dataset, which outperforms the state-of-the-art accuracy levels. There is also development of a strong feature extraction, fusion, and classification pipeline which would be helpful in increasing the performance of FER in difficult settings like occlusion, variability in poses and pixel density.This work not only advances FER methodologies but also provides valuable insights for future research aimed at improving non-verbal communication systems.

The primary results of this research are as follows:

- A novel hybrid approach fusing handcrafted features (HoG, Facial Landmarks) with deep learning features (CNNs) to improve FER accuracy under occlusions.
- Enhances FER robustness against occlusions (masks, scarves, sunglasses) for real-world applications.
- Comparative analysis proves feature fusion outperforms individual handcrafted and deep learning methods.
- Achieves 96% accuracy on RAF-DB and Affect-Net, surpassing existing FER techniques.
- Establishes a scalable feature extraction, fusion, and classification pipeline adaptable to diverse conditions.
- Enables applications in HCI, AI-driven emotion analysis, education, psychology, and security.

This research contributes to advancing occluded FER methodologies by integrating handcrafted and deep learning-based features, offering a scalable and efficient solution for real-world emotion recognition challenges.

In the subsequent sections, Section 2 presents a review of related works. Section 3 outlines the proposed model architecture and provides an overview of the dataset utilized in this study. Section 4 presents the experimental results and visualizes the model's performance. Lastly, Section 5 concludes the paper.

RELATED WORK

It has been proved that facial expression representation is not only influenced by muscular deformation of facial structure, most of the studies in the area of facial expression recognition research have investigated only the effects of muscular deformation of facial structure. Shi et al. [9] proposed an Amend Representation Module (ARM). Convolution padding erodes the feature map while also aiding in the acquisition of edge information. ARM serves as a replacement for the pooling layer. To handle Padding Erosion, it can be integrated into the back end of any network. The validation accuracy for RAF-DB is 90.42%, Affect-Net is 65.2%, and SFEW is 58.71%, respectively. Poux et al. [10] offered a method utilizing an auto-encoder architecture has been proposed. This method's three basic steps make up the full process. The first step is to calculate the optical fluxes between each frame of a sequence of occluded faces. The second stage entails utilizing a learned auto-encoder to recover optical flows that were distorted by the occlusion. The auto-encoder is trained using pairs of occluded and non-occluded optical fluxes. The classification stage of predicting the expression then uses the recovered optical flows directly.

A CNN architecture that has been trained to recognize facial expressions serves as the foundation for evaluating facial expression recognition. Our strategy is assessed using the CK+ dataset. Yong Xuan et al. [11] proposed Convolution neural network with attention mechanism, please (ACNN). Based on specific areas of the face, humans can identify various facial expressions. When certain areas of the face are blocked but others are not Automatic facial patch blocking is detected by ACNN, which focuses mostly on informative and unblocked patches. He presented Patch-based ACNN (pACNN) and global-local-based ACNN are the two variations of ACNN, for various face regions of interest (gACNN). Only regional face patches are considered by pACNN. Through the use of gACNN, the local representations at the patch level are combined with the global representation at the image level. The datasets RAF-Db, AffectNet, and FED RO are used. Houshmand et al. [12] proposed a geometric model that may be used with current FER datasets to recreate the occlusion brought on by a Samsung Gear VR headset.

The networks were then further fine-tuned on the FER+ and RAF-DB datasets using the Starting with two pre-trained networks, VGG and ResNet, the transfer learning approach [13]. According to experimental findings, this method produces outcomes that are equivalent to those of other approaches in which the use of a cheap VR headset causes occlusion. Acsintoae et al. put forth a three-part procedure [14] where first, he uses the traditional teacher-student training method, when the teacher is a CNN trained on clearly visible faces, and CNN, which is trained to identify hidden faces, is the pupil.

Next, we provide a novel triplet loss-based strategy for knowledge distillation. The goal of training is to reduce the gap between a student's anchor embedding on CNN, which takes input from obscured faces, and a teacher's positive embedding on CNN, which was trained on fully visible faces, smaller than the distance between the CNN negative embed and the student which employs a class distinct from the anchors. He runs tests on the benchmarks FER+ and AffectNet using the CNN architectures VGG-f and VGG-face, demonstrating that knowledge distillation can perform much better than current techniques for occluded faces in VR [15]. Kai Wang et al. [16] developed a method to capture the importance of face areas for occlusion and position variant FER in an adaptive manner, A brand-new Region

Attention Network (RAN) is what we suggest. a fundamental convolutional neural network produces a variety of area features, which the RAN aggregates and embeds into a short, fixed-length representation. He suggests a region-biased loss in the last phase to encourage higher weights for the most important areas. He tested his RAN and region-biased loss on four well-known datasets, including FERPlus, AffectNet, RAF-DB, and SFEW, in addition to our test datasets, Yong Li al [16] proposed a technique to automatically concentrate on the most discriminatory un-occluded regions while perceiving the occluded section of the face, we present an end-to-end trainable Patch-Gated Convolution Neutral Network (PG-CNN). The PG-CNN divides a map of intermediate features numerous patches based on the locations of relevant facial features in order to locate prospective interest areas on the face.

Then, using PG-CNN reweights each patch in accordance with its significance or degree of unobstructedness as assessed by the patch itself, as recommended by the Patch-Gated Unit. On the AffectNet and RAF-DB datasets, an experimental investigation was carried out to confirm the efficacy of the suggested strategy [18], [19]. Dan Zeng et al. [20] proposed an occlusion mask adapter was created as a bridge in simultaneous occlusion invariant deep networks because the corrupted features induced by occlusion can be found within an occlusion segmentation (SOIDN). Use deep CNN to extract occlusion-resistant traits. Coherently combine the occlusion segmentation network with the face recognition network and optimize in a concurrent architecture occlusion invariant feature to learn. The training dataset consists of synthetic occluded CASIA-Webface and CASIA-Webface [21]. Levi & Hassner, 2015) before training the architecture image is fine tune (to overcome illumination problems) using local binary patterns for emotion estimation as shown in figure 2.

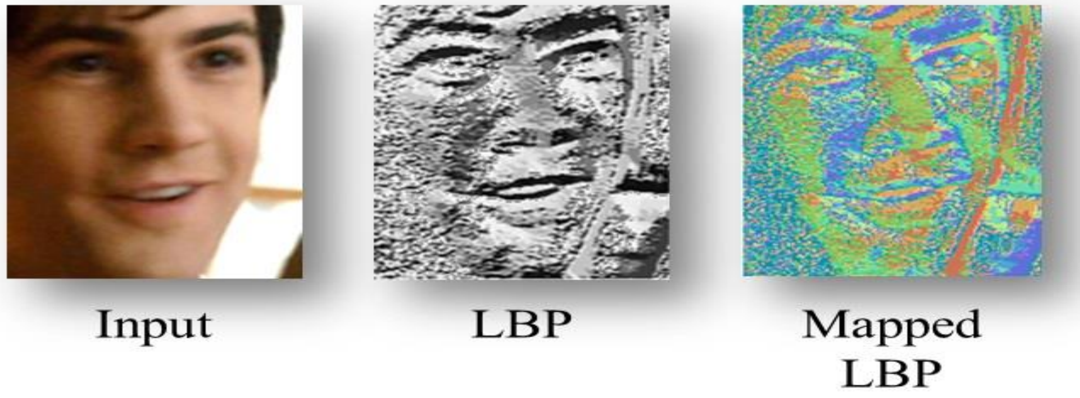


Figure 2.
Gil Levi, Hassner, LBP Mapped Image [21]

Then RGB original image and LBP mapped code is used for different CNN architectures (VGG-S, VGG-2048, VGG-M-4096) individual training as well as ensemble architectures training (VGG-S-LBP 10 Cyclic, VGG-M-4096-LBP 1). King et al. [22] Ensemble architectures results were good and the margin was least.

Egede et al. [23] handcrafted and learned features are used for facial pain estimation. Histogram of gradients, geometric features and deep CNN features ensemble using SVR (support vector regression) for estimating pain intensity. For comparison RMSE (root mean square error) and CORR (Pearson Correlation) is used. Proposed method outperforms. 14% is increased in performance measure. 2%

increase in CORR and 70% reduction in RMSE. Georgescu et al. [24] different CNN architectures are combined with handcrafted features using BOVW (Bag of Visual Words) to achieve high accuracy results for facial expression recognition. Three CNN architectures VGG-13, VGG-f, VGG-Face used. VGG-13 trained from scratch only others were used pre-trained. SIFT --- KNN --- SVM used for handcrafted features. Proposed architectures out performed by least margin as compared to recent proposed approaches shown in figure 3 This happens normally for each human in the course of their life. The issues it causes are the facial structure, which changes marginally all through the early years. Related to maturing, the skin surface gets more unpleasant with lines predominantly on the brow and close to the eyes. For the most part brought about by lighting conditions when a picture is caught.

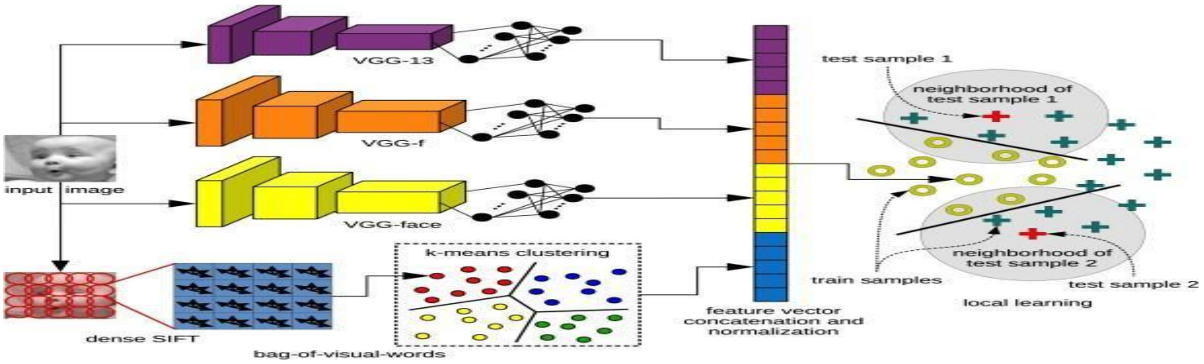


Figure 3.

Georgescu, Ionescu, Occlusion of 3 VGG Net and Bag of Word [24]

In Fan et al. [25] worked on MRE-CNN (Multi region Ensemble –CNN) was proposed for better facial expression estimation. Alextnet and VGG-Net was used. Firstly, system detected face, then divided it into different part like mouth, nose, ear etc. Then softmax was used as activation function, then ensemble features. Proposed architecture results were better as compared to previously proposed solutions. In Alshazly et al. [26] handcrafted features ensembled with CNN features to recognize ear. Handcrafted approaches were used like LPQ (Local phase quantization feature), Local binary patterns, Histogram of gradients, BSIF (Binarized statistical image features), POEM (Pattern of oriented edge magnitude). For learned features CNN was used. All possible combination of handcrafted with handcrafted and handcrafted with learned features were used for facial estimation.

Combined handcrafted features with CNN results were better. Normalization of handcrafted features results were better. Ortega et al. [27] audio (Handcrafted) and video features is used for emotion detection. Two types of fusion are used. Early fusion and Late fusion. Different combination proposed like CNN + Transfer learning, CNN + Pre-Processing, SVM + Audio features. Proposed approach used pre-train CNN features combined with handcrafted audio feature outperforms. This present work's essential center is to make a Deep Convolutional Neural Network (DCNN) model that orders five diverse human facial feelings.

The study explores the impact of various parameters on face recognition performance using HOG descriptor settings, such as window size and cell size. It discusses advancements in facial expression recognition, including face detection, feature extraction, and ethnic demeanor analysis, as well as emotion changes through video sequences. The research also examines the accuracy of different datasets used for training and testing, considering factors like static vs. dynamic conditions and lab vs. non-lab environments.

The study examines the precision of different algorithms in emotion recognition, highlighting the superior accuracy of MHL sensors over webcams for detecting subtle micro-expressions. Testing with a broad database for real-time emotion recognition demonstrated the general applicability of the HOG feature extraction method across various datasets. The study also did its share of optimization of HOG parameters in face recognition, such as gamma light correction, spatial angle and image resolution in adverse conditions, by utilizing FERET database.The study compares dense grid HOG with two local facial feature extraction techniques, Gabor wavelets and Local Binary Patterns (LBP), for face recognition. HOG features are extracted from non-overlapping dense grid face images, with performance analyzed under various conditions. The system was tested on side-view face images from the CMU-Multi PIE database, exploring applications in security.

The work also discusses the advancements in facial expression recognition, including emotion analysis, pattern recognition, and ethnic demeanor recognition, highlighting recent developments by Wu et al. in ethnic expression recognition. Targets another technique presented in an examination for Facial demeanor acknowledgment. It utilizes the FER2013 information base comprising seven classes (Surprise, Fear, Angry, Neutral, Sad, Disgust, Happy). In the previous few decades, exploring strategies to perceive outward appearances has been dynamic exploration territory, and numerous applications have created for highlight extraction and derivation. In any case, it is as yet testing because of the high-infraclass variety.

Existing investigations on FEA under halfway impediment need to complete benchmark datasets. It incorporates a thick arrangement of different sorts of successive average facial restrictions. A very much commented on ground certainties of outward appearances by discrete classifications as well as AUs and dimensional tomahawks. Working in planning face impediment identification methods, the dependably decide the precise boundaries of facial impediments. Future FEA frameworks improve by oh et al. [29] in dealing with facial impediment expected to extend from misleadingly forced to normally happening limitation; 2D to 3D face information bases; manual face preprocessing to programmed impediment recognition and incorporation, static 2D dim to worldly 3D shading highlights; a solitary looks to numerous countenances of a gathering of individuals, a solitary face methodology to various sound, visual and physiological modalities, shallow engineering to additional inside and out and more extensive models, prototypical feelings to AU-coded, persistently spoke to feelings and miniature articulations.

In Gao et al . [30] multi stream CNN features figure 1.5 that show the combination was used for driving behavior recognition. Abnormal driving can cause different swear accidents. To overcome this problem stated architecture was proposed. It gets input from different streams of CNN architecture. Combined features, using SVM (Support Vector Machine) late fusion. For early fusion score-based approach is used as shown in figure below. As you can see from the picture there are three different streams of different feature extraction techniques we layer combined to one before classification. Approach compared with different handcrafted features. Early fusion using score base results were batter as compared to late fusion using SVM. In Liu et al. [31] MLVSF (Multi-layer visual feature fusion) approach was proposed. In medical image analysis handcrafted feature were performing well as compared to automated features. Proposed approach was used to combine different handcrafted, mid-level and dense level features. These features extracted by LBP, Bag of Visual Words and CNN (Alexnet, VGG Net) respectively. Experimental results showed that MLVSF can improve CNN for better accuracy. The comprehensive details of the literature as shown in table 1.

Table 1.
Literature Review

Study	Dataset	Model	ACC
Jiawei Shi et al. (2021) [32]	RAF-DB, AffectNet, SFEW	Multi-Layer Visual Feature Fusion (MLVSF)	90.42% (RAF-DB), 65.2% (AffectNet), 58.71% (SFEW)
Delphine Pouxet al.(2020) [33]	CK+	Auto Encoder method	Recognition improved
Yong Li et al.(2022) [34]	RAF-DB, AffectNet, FED RO	CNN	98%
Egede et al. (2021) [35]	Custom dataset	ANN	14% performance, 2% CORR
Georgescu et al.(2022) [36]	VGG-Face, FERET, Others	CNN	70%
Hansley et al. (2024) [37]	Custom dataset	CNN	80%
D. Liu et al. (2019) [38]	Custom dataset	visual feature fusion (MLVSF)	78%

METHODOLGY

The present study focuses on comparing different combinations of Histogram of Gradients (HoG), facial landmarks, HoG features derived from a sliding window, and CNN features for facial emotion estimation. Four distinct combinations were tested: (1) using CNN for both feature extraction and classification, (2) combining CNN with HoG features, (3) integrating CNN, facial landmarks, and HoG features, and (4) utilizing CNN, sliding window features, facial landmarks, and HoG features. For every iteration, classification was performed using the Fully Connected (FC) layer of the CNN. The results demonstrated that the inclusion of more features in the feature vector led to an increase in accuracy. Figure 4 illustrates this joint learning approach for racial identity-aware facial expression recognition.

The experiments were conducted in three phases: first, with 1000 images per label; second, with 5000 images per label; and finally, using the entire dataset. Interestingly, as the number of images per label increased, the accuracy decreased.per label increased, the accuracy decreased. This was attributed to the need for more computational power and advanced graphics processing units (GPUs), which resulted in overflow issues and a subsequent drop In accuracy.The subsequent chapter delves into the detailed implementation procedure, including the requirements necessary to fully understand the process. Fig.4 illustrate the proposed methodological in this study. The methodology for building and preparing the CNN included several steps. First, the testing and training datasets were prepared. Then, the CNN layers were constructed using the TensorFlow library,followed by the selection of an appropriate optimizer. Next, the feature sets were merged, and the network was trained, with checkpoints saved at regular intervals. Finally, the trained model was tested to evaluate its performance.

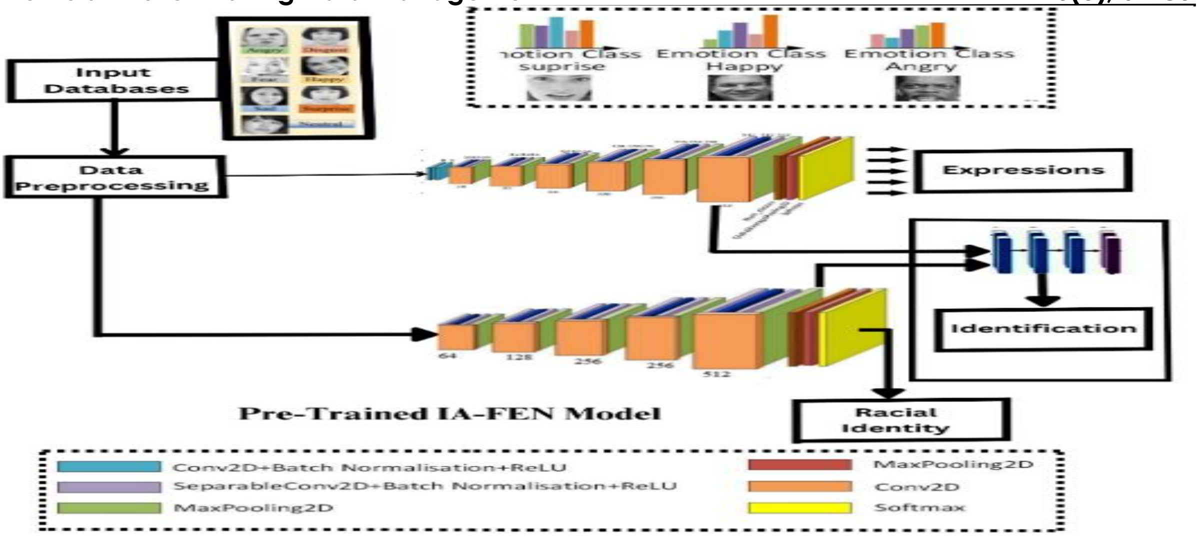


Figure 4.
Proposed Pre-Trained IA-FEN Model for Facial Expression Recognition

The methodology for building and preparing the CNN included several steps. First, the testing and training datasets were prepared. Then, the CNN layers were constructed using the TensorFlow library, followed by the selection of an appropriate optimizer. Next, the feature sets were merged, and the network was trained, with checkpoints saved at regular intervals. Finally, the trained model was tested to evaluate its performance.

Input Databases

The database used to test and evaluate is called FER 2013, and contains a total amount of 28,709 grayscale images of dimensions 48 x 48 pixels, as can be seen in Table.2. The database is split into two major parts a training set and a testing set. The training set comprises 28,709 images, while the testing set is further subdivided into two subsets: the private test set, containing 28,709 images, and the public test set, also containing 28,709 images. The model has a validation set as well, and in Table.3 the validation set is represented. The annotation of the dataset uses emotion labels, which sets an inclusive basis of training, validation, and testing in emotion recognition applications.

Table 1.
Classes of FER Dataset

Dataset Classes	Images
Neutral	4965
Anger	3995
Disgust	436
Fear	4097
Happy	7215
Sadness	4830
Surprise	3171
Anger	3995
Resolution	48 * 48
Total	28709

There is 28709 training image in FER 2013. It also has a validation and test set. Here is the emotion label list of those.

Table 2.
Validation and Testing Dataset Details

Name	Validation Set	Testing Set
Total	3589	3589
Neutral	607	636
Anger	467	491
Disgust	56	55
Fear	496	455
Happy	895	879
Sadness	653	653
Surprise	415	416
Anger	415	415

Data Augmentation

With the labelled original dataset, synthetic images can be created by various transformations to the original images. Image Data Generators is one transformations method used for generating more training data from the original data to avoid model overfitting. There are various data augmentation techniques. The selected data augmentation explains in Table 4. These techniques were: flipping images horizontally or vertically, rotating images at 40 degrees, rescaling outward or inward, randomly cropping, translating by width and height shifts, whitening, shearing, zooming and adding Gaussian noises to prevent model overfitting and enhance learning capability. Some example images generated by using the Image Data Generator are shown in Table 4.

Table 3.
Data Augmentation Settings

Transformations Ranged From	Setting
Scale transformation	Ranged from 0 to 1
Rotation transformation	25 ◦
Zoom transformation	0.2
Horizontal flip	True
True Shear transformation	20 ◦

By augmenting the dataset with these modified versions of the original data samples, machine learning models can learn to generalize better and become more resistant to overfitting. Additionally, data augmentation can help address issues related to imbalanced datasets by generating synthetic examples of minority classes. Data augmentation is a powerful tool for improving the performance and robustness of machine learning models, particularly in scenarios where collecting additional labeled data may be costly or impractical that are represented in Table 5.

Table 4.
Image Augmentation Techniques

Name of the Dataset Classes	After Augmentation Validation Test	After Augmentation Testing Sett
Neutral	1214	1252
Anger	934	982
Disgust	112	110
Fear	992	1052
Happy	1790	1758

Sadness	1306	1188
Surprise	830	832
Anger	830	830

Dataset Processing

The FER 2013 dataset is provided in CSV format, containing columns for emotion labels, pixel values, and usage. The "Usage" column indicates how each image is intended to be used within the dataset. The entries labeled "Training" are designated as the training dataset, "Public Test" is used as the validation set, and "Private Test" is reserved for the testing set. This structure allows for clear separation of data for training, validation, and testing during the model evaluation process. The facial appearance is detected, cropped, and aligned using a 68-point facial landmark. Landmark vector of shape (68,2) saves for dataset. To remove the difference with emotion is changed where 0,1,2, 3, 4, 5, and 6 is equal to Neutral, Anger, Disgust, Fear, Happiness, Sadness, and Surprise respectively.

Model Identity-Aware Facial Expression Network (IA-FEN)

The process for Model 1(IA-FEN) as shown in figure 4 began with obtaining the raw data. An optimization algorithm was run to enhance performance, and the parameters were updated in the parameter file. After doing this optimized set up, the model was trained. After training, CNN features were extracted and these features were applied on the images to classify the images using CNN model. Lastly, the performance of the model was checked to know whether accuracy and effectiveness of the model is good.

The proposed elegant approach to learning facially expressed features, introduces the technology that learns expression features in images simultaneously with learning attributes of the racial identity in a pre-learned identity network. These characteristics of racial identity add more context to the system as they can recognize the expression of various cultures better. The IA-FEN integrates the identification of the same across racial groups using facial features and race identity in a single multi-cultural facial expression (MCFE) by including racial identity and facial expression features into a single-representation. This has an added advantage of enhancing more robust and accurate expression analysis. This combined set of features is further fed on the fully connected layers of the IA-FEN framework giving rise to a more comprehensive and more sophisticated recognition system.

This module isolates complex facial structures in pictures fed into this module. A deep residual learning uses shortcut connections to optimize performance whereas batch normalization layers ensure faster convergence and precision. Average pooling is used instead of flattening feature maps using fully connected layers, producing a small and a more efficient representation that has fewer parameters. After that, a fully connected layer combines racial identity and facial expression features. The upper and lower network parts combined will create the whole IA-FEN framework. Throughout the training, the concatenated facial expression features combined with racial identity features form a feature vector which in turn is fed-in to the fully connected layers so that IA-FEN is successfully able to interpret and differentiate among multi-cultured facial expressive differences.

One of the novelties of the IA-FEN model consists in incorporating residual blocks into the ResNet framework and optimizing the performance on different recognition tasks. ResNet relies on a deep residual learning framework, in which learning to refine residual mappings is more useful than attempting to learn such as to directly optimize over original feature mappings. It is the ResNet architecture that makes it especially appropriate to use when collecting very subtle emotional characteristics. However deferring neural networks become more difficult, and this results in the performance degradation or increased training errors as depth increases. To this end, skip connections are forethought in residual blocks to keep important input information intact and to propagate it steadily to additional layers.

The input is x and a desired mapping to be learned is f(x) as shown in Figure 5. The traditional method in the left side of the figure suggests that the model learns f (x) directly. On the right, another alternative is using the residual block architecture in which the model is tasked with learning the residual f(x) x. The presence of original input x in learned mapping subsumes reconstruction of f(x) better. The figure is depicting a residual or shortcut connection using solid line that guarantees flow of uninterrupted information and avoids the damaging of signals. Such a framework increases the stability of deep learning training, gradient flow and the overall effectiveness of the model, which shows the success of the residual block in the deep neural networks.

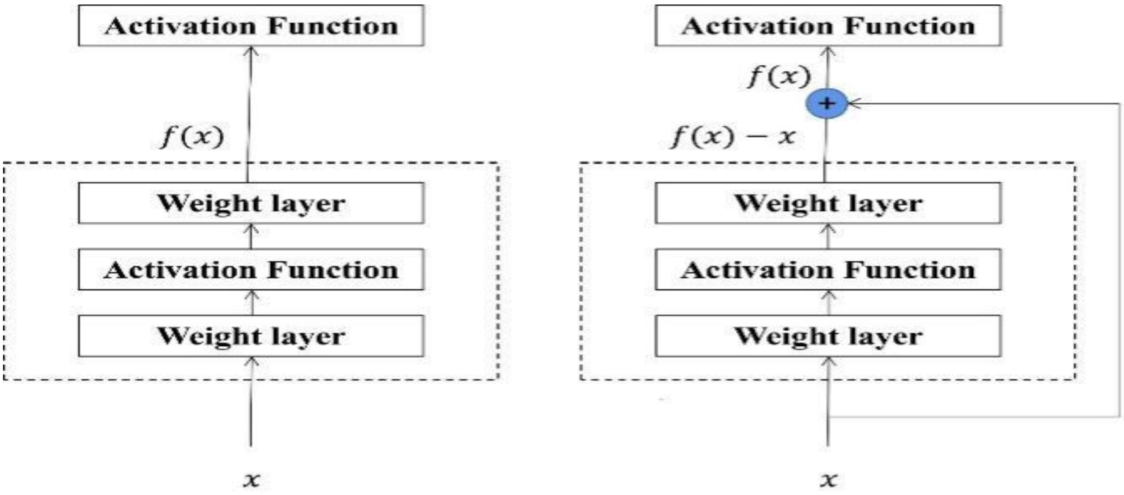


Figure. 5.
Mapping Model Activation Function

Evolution Criteria

Performance of the proposed method was evaluated with the key evaluation measures, such as the accuracy, precision, recall, and F1 score. Further, to have more detailed analysis of the models, confusion matrix was applied. Though accuracy is easy to interpret and a simple measure, it can be used only on datasets with equal proportion of all classes. Accuracy is an unreliable advantage in the situation of class imbalance. In order to get a proper assessment, additional measures like precision, recall, and F1 score were also used which gave a more comprehensive and precise picture of how the models performed.

RESULTS

The prior chapter presented models and each of them aimed at solving the issue of occluded facial expression recognition. These models adapted diverse features extraction methods, including Facial Landmarks, Histogram of Gradients (HoG), HoG with Sliding Windows, and CNN-extracted features. Ultimately, all models utilized a Convolutional Neural Network (CNN) for classification. This chapter focuses on evaluating the accuracy of these models, analyzing their performance, and conducting a comparative analysis to assess the effectiveness of the combined features and classification strategies employed.

Experimental Results without Augmentation

In this experiment results for our proposed technique using datasets firstly, FER 2013 is used for experimentation. This information dataset comprises of the 7 essential articulations which were presented by 10 female models. 716 images are total in this dataset, angry images are 30, disgust 29 images, fear images are 32, happy images are 31, contain 30 neutral, 31 sad images, and 30 are of surprise. Separately every model gave roughly three pictures to every apparent expression. To each picture was spared in grayscale with a goal of 256 × 256. disgust is 59, 25 images are fear, 69 images of happy, 28 images with sad expressions, and surprise 83 images. Experiments done using KDEF dataset produces very good results having accuracy of 69% in angry expression as shown in table 6.

Table 5.
Evaluation parameters for all combined datasets

Label	Precision	f1 score	MCC
Anger	0.66	0.63	0.62
Contempt	0.64	0.61	0.61
Disgust	0.52	0.54	0.5
Fear	0.68	0.64	0.63
Happy	0.56	0.58	0.56
Sadness	0.59	0.58	0.56
Surprise	0.59	0.63	0.61
Anger	0.66	0.63	0.62

Proposed Results of IA-FEN model with Augmentation

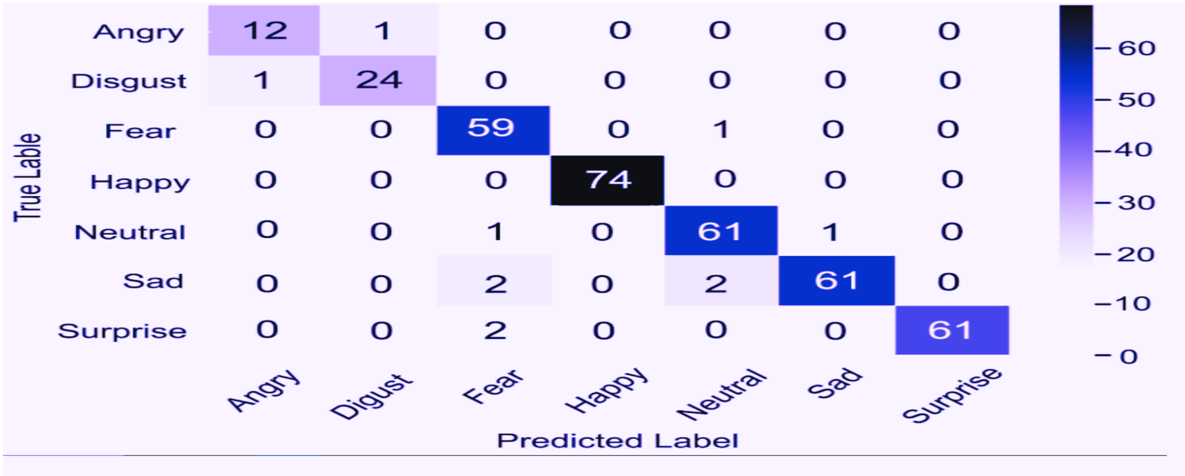
The first experiment evaluates the performance of four models using a balanced dataset of 1,000 images per label, with labels containing fewer than 1,000 images utilizing all available data. The test accuracies achieved by the models highlight the progressive improvement in performance as additional feature extraction techniques are integrated. The experimental analysis shows that the IA-FEN Facial Expression Recognition system achieves a maximum accuracy of 0.9697. Table 7 outlines the results for each facial expression based on precision, recall, and F1 score. The table reveals that the emotions anger, sadness, and fear are most prone to misclassification. Furthermore, the results indicate that happiness and surprise are more easily recognized compared to sadness, fear, and anger. The highest recognition accuracy (100%) is achieved for happy, while the lowest (92.00%) is recorded for angry. These results demonstrate the effectiveness of integrating diverse features to enhance facial expression recognition, particularly in challenging scenarios involving occlusion.

Table 6.
Experiment results on proposed method

Label	Precision	Recall	F1 Score
Anger	0.92	0.92	0.92
Contempt	0. 96	0. 96	0. 96
Disgust	0. 92	0. 98	0. 95
Fear	0.92	0.98	0.95
Happy	1.00	1.00	1.00
Sadness	0.98	0.98	0.85
Surprise	0.59	0.97	0.98
Anger	0.66	0.63	0.62

Confusion matrix

The performance of the IA-FEN model is visualized in Figure.6 using a confusion matrix for seven facial expressions: angry, disgust, fear, happy, neutral, sad, and surprise.



True Lable	Angry	Digust	Fear	Happy	Neutral	Sad	Surprise
Angry	12	1	0	0	0	0	0
Disgust	1	24	0	0	0	0	0
Fear	0	0	59	0	1	0	0
Happy	0	0	0	74	0	0	0
Neutral	0	0	1	0	61	1	0
Sad	0	0	2	0	2	61	0
Surprise	0	0	2	0	0	0	61

Figure 6.
Confusion matrix of IA-FEN Model

The matrix provides a detailed breakdown of each emotion's classification. The y-axis indicates the predicted emotion, and the x-axis indicates the true emotion. Diagonal entries represent correct classifications, while off-diagonal entries represent errors. Each row in the matrix highlights the confusion between specific pairs of emotions. Training the model with the entire dataset resulted in the accuracy and loss curves depicted in Figure 7. The IA-FE Network achieved its lowest error rate after completing 300 epochs. Figure 6b further visualizes the accuracy curve for both training and testing phases on the dataset.

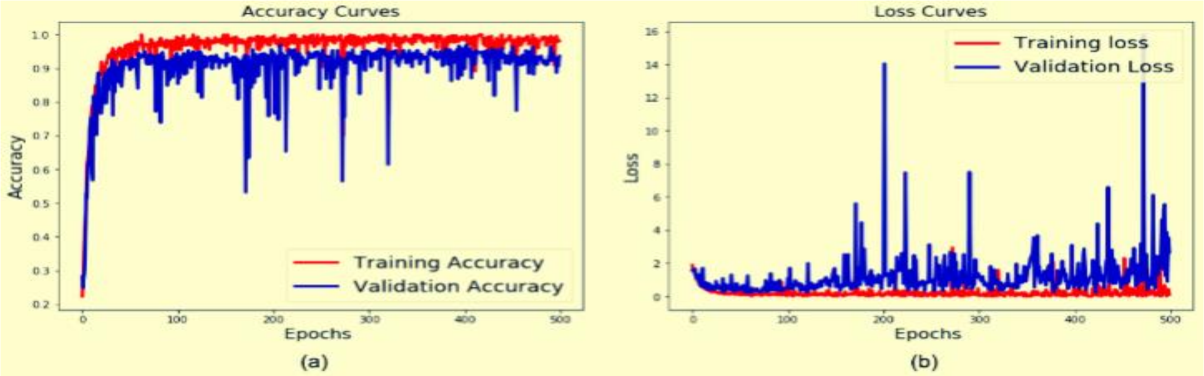


Figure. 7.
Accuracy and Loss Curve of IA-FEN Model

DISCUSSION AND ANALYSIS

The study compares a Deep Conventional Neural Network for automatic Occluded Facial Expression Recognition deep learning model (IA-FEN) achieved 96.97% accuracy. Results confirm that facial expression representation varies across cultures, affecting recognition accuracy and learning algorithm performance. Happy and surprise expressions were the most accurately recognized, while sad and fear were the hardest to classify due to inter-expression resemblance. The study found that angry expressions were the most confusing due to muscle distortions affecting recognition accuracy.

Comparison with State-of-the-Art Models

The comparison with existing approaches highlights the limited research on multicultural facial expression recognition using combined datasets. Previous works table 5, such as Zhang et al. [41], applied Deep Convolutional Neural Networks to a multicultural dataset, achieving 94.12% accuracy. Similarly, Liu et al. [42] employed Hybrid Feature Fusion with CNNs on the Multicultural Faces Dataset, reaching 92.56%, while Kim et al. [43] used Transfer Learning with ResNet, obtaining 91.85%. Viola Jones et al. (2021) utilized Convolutional Neural Networks (CNNs) for facial expression recognition, employing facial landmarks extracted from benchmark datasets JAFFE, MUG, CK, and MMI, achieving recognition accuracies of 84%, 89.19%, 85.42%, and 84.33%, respectively.

Their method revealed the opportunity of CNNs to deal with different facial expressions on various datasets. Eigen Vector et al. (2019) examined the Multi-Layer Perceptron (MLP), the Support Vector Machines (SVM), and the J48 decision tree in faculty expression classification. In their assessment, SVM attained an accurate rate of 50 percent, whereas J48 proved to be more accurate at 70 percent, which goes to show traditional machine learning limitations when compared to deep learning algorithm in facial expression recognition practices.

The techniques were mostly based on handcrafted features and ensembles. By contrast, IA-FEN model proposed in the research, operating on a FER 2013 dataset, combines the tasks of feature extraction and classification in a single framework and has the accuracy of 96.97% as displayed in Table 8. Adding racial identity characteristics, a specific feature of IA-FEN successfully reduces cultural, structure of the face, and expriation diversities, which renders it more weighty than the alternatives available.

Table 8.
Comparison with previous studies

Author	Methodology	Dataset & Accuracy
Zhang et al. (2023) [41]	Deep Convolutional Neural Networks	Multicultural Dataset - 94.12%
Liu et al. (2022) [42]	Hybrid Feature Fusion with CNNs	Multicultural Faces Dataset - 92.56%
Kim et al. (2021) [43]	Transfer Learning with ResNet	Multicultural Facial Expressions Dataset - 91.85%
Viola Jones et al. (2021) [44]	CNN	Facial Landmarks JAFFE, MUG, CK, MMI (84%, 89.19%85.42%84.33%)
Eigen Vector et al. (2019) [45]	MLP, SVM and J48 decision tree	SVM 50% and J48 decision tree rate is 70%.
Proposed IA-FEN	FER 2013	96.97%

CONCLUSION

This study introduces an enhanced occlusion-based facial emotion recognition technique that effectively combines handcrafted and deep learning features. By integrating CNN features, HoG features, and facial landmarks into a unified feature vector, the proposed approach improves recognition accuracy, particularly in scenarios with limited datasets. Comparative analysis with existing methods highlights the superiority of our models in handling occluded expressions, demonstrating robustness and adaptability. While the model achieved a high accuracy of 96% on benchmark datasets (RAF-DB and Affect-Net), performance slightly declined with larger datasets. These findings underscore the potential of hybrid feature fusion techniques in advancing facial emotion recognition, offering promising applications in human-computer interaction, psychological studies, and security systems. Future research can explore optimizing feature selection strategies and extending the approach to more diverse and large-scale datasets.

LIMITATIONS

Contrarily, one notable limitation has also been noted when employing occlusion-based techniques, namely that the noise or disturbance level has increased dramatically when using larger data sets. Finding the best methods and filters to lessen the ambiguity caused by noise, disturbance, and overfitting will be the job of the future, while enhancement and restoration offer better viewing than the methods currently in use. Emerging upside-down picture resolution techniques and system pass are crucial to adapt to and expand upon in the methods developed, which can substantially aid in analyzing the data.

FUTURE WORK

The proposed approach improves execution time through parallelization, leveraging independent image divisions to reduce overfitting. It enhances occlusion-based methods, performing well on diverse images with limited data. Some overcorrections caused minor information loss, but the method outperformed existing techniques. Future versions will address these limitations for further refinement.

DECLARATIONS

Acknowledgement: We appreciate the generous support from all the contributor of research and their different affiliations.

Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.

Availability of data and material: In the approach, the data sources for the variables are stated.

Authors' contributions: Each author participated equally to the creation of this work.

Conflicts of Interests: The authors declare no conflict of interest.

Consent to Participate: Yes

Consent for publication and Ethical approval: Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their consent.

REFERENCES

A. Bilal *et al.*, “Improved support vector machine based on CNN-SVD for vision-threatening diabetic retinopathy detection and classification,” *PLoS One*, vol. 19, no. 1, p. e0295951, 2024, doi: 10.1371/journal.pone.0295951.

A. R. Khan, “Facial emotion recognition using conventional machine learning and deep learning methods: Current achievements, analysis and remaining challenges,” *Information*, vol. 13, no. 6, p. 268, 2022, doi: 10.3390/info13060268.

Abaidullah, A., & Basheer, M. F. (2024). Nexus among Entrepreneurial Activities, Human Capital, and Economic Growth to achieve Sustainable Development Goals (SDGs): Moderating Role of Financial Development. Journal of Finance and Accounting Research, 6(1), 1-27.

Abbasi, M. D., Sajid, Z., Khawer, S. K., & Mir, S. Z. (2025). Automatic Speech Recognition by Using Neural Network Based on Mel Frequency Cepstral Coefficient.

Aslam, S., Usman Javeed, M. ., Maria Aslam, S. ., Iqbal, M. M., Ahmad, H. ., & Tariq, A. . (2025). Personality Prediction of the Users Based on Tweets through Machine Learning

Techniques. Journal of Computing & Biomedical Informatics, 8(02). Retrieved from https://www.jcbi.org/index.php/Main/article/view/796.

B. Houshmand and N. M. Khan, “Facial expression recognition under partial occlusion from virtual reality headsets based on transfer learning,” in *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*, 2020, pp. 70–75, doi: 10.1109/BigMM50055.2020.00021.

Basheer, M. F., Sabir, S. A., & Hassan, S. G. (2024). Financial development, globalization, energy consumption, and environmental quality: Does control of corruption matter in South Asian countries?. Economic Change and Restructuring, 57(3), 112.

C. Sirithunge, A. G. B. P. Jayasekara, and D. P. Chandima, “Proactive robots with the perception of nonverbal human behavior: A review,” *IEEE Access*, vol. 7, pp. 77308–77327, 2019, doi: 10.1109/ACCESS.2019.2922054.

D. Liu, Y. Liu, S. Li, W. Li, and L. Wang, “Fusion of handcrafted and deep features for medical image classification,” in *J. Phys.: Conf. Ser.*, vol. 1345, no. 2, p. 022052, 2019, doi: 10.1088/1742-6596/1345/2/022052.

D. Poux *et al.*, “Facial expressions analysis under occlusions based on specificities of facial motion propagation,” *Multimedia Tools Appl.*, vol. 80, pp. 22405–22427, 2021, doi: 10.1007/s11042-021-11050-1.

E.-G. Lee, I. Lee, and S.-B. Yoo, “ClueCatcher: Catching domain-wise independent clues for deepfake detection,” *Mathematics*, vol. 11, no. 18, p. 3952, 2023, doi: 10.3390/math11183952.

G. Devasena and V. Vidhya, “A study of various algorithms for facial expression recognition: A review,” in *2021 International Conference on Computational Intelligence and Computing Applications (ICCICA)*, 2021, pp. 1–8, doi: 10.1109/ICCICA.2021.00012.

G. Levi and T. Hassner, “Emotion recognition in the wild via convolutional neural networks and mapped binary patterns,” in *Proc. 2015 ACM Int. Conf. Multimodal Interact.*, 2015, pp. 503–510, doi: 10.1145/2818346.2830593.

H. A. Amirkolaee, D. O. Bokov, and H. Sharma, “Development of a GAN architecture based on integrating global and local information for paired and unpaired medical image translation,” *Expert Syst. Appl.*, vol. 203, p. 117421, 2022, doi: 10.1016/j.eswa.2022.117421.

Huo, S., Ni, L., Basheer, M. F., Al-Aiban, K. M., & Hassan, S. G. (2024). The role of fintech, mineral resource abundance, green energy and financial inclusion on ecological footprint in E7 countries: New insight from panel nonlinear ARDL cointegration approach. Resources Policy, 94, 105083.

J. Anil *et al.*, “Literature survey on face recognition of occluded faces,” in *2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT)*, vol. 1, 2024, pp. 1930–1937, doi: 10.1109/ICCPCT.2024.00056.

J. Anil *et al.*, “Literature survey on face recognition of occluded faces,” in *2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT)*, vol. 1, 2024, pp. 1930–1937, doi: 10.1109/ICCPCT.2024.00056. [26] H. Alshazly, C. Linse, E. Barth, and T. Martinetz, “Handcrafted versus CNN features for ear recognition,” *Symmetry*, vol. 11, no. 12, p. 1493, 2019.

J. D. S. Ortega, P. Cardinal, and A. L. Koerich, “Emotion recognition using fusion of audio and video features,” in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2019, pp. 3847–3852, doi: 10.1109/SMC.2019.8914663.

J. E. T. Akinsola, O. Awodele, S. O. Kuyoro, and F. A. Kasali, “Performance evaluation of supervised machine learning algorithms using multi-criteria decision making techniques,” in *Proc. Int. Conf. Inf. Technol. Educ. Dev. (ITED)*, 2019, pp

J. Gao, J. Yi, and Y. L. Murphey, “Multi-scale space-time transformer for driving behavior detection,” *Multimedia Tools Appl.*, vol. 82, no. 16, pp. 24289–24308, 2023, doi: 10.1007/s11042-023-15129-5.

J. Shi, S. Zhu, and Z. Liang, “Learning to amend facial expression representation via de-albino and affinity,” *arXiv preprint arXiv:2103.10189*, 2021.

J. Shi, S. Zhu, D. Wang, and Z. Liang, “ARM: A lightweight module to amend facial expression representation,” *Signal Image Video Process.*, vol. 17, no. 4, pp. 1315–1323, 2023, doi: 10.1007/s11760-023-02352-x.

J.-J. Liu, Q. Hou, and M.-M. Cheng, “Dynamic feature integration for simultaneous detection of salient object, edge, and skeleton,” *IEEE Trans. Image Process.*, vol. 29, pp. 8652–8667, 2020, doi: 10.1109/TIP.2020.3020789.

Javeed, M. U., Shafqat Maria Aslam, Hafiza Ayesha Sadiqa, Ali Raza, Muhammad Munawar Iqbal, & Misbah Akram. (2025). Phishing Website URL Detection Using a Hybrid Machine Learning Approach. Journal of Computing & Biomedical Informatics. Retrieved from https://jcbi.org/index.php/Main/article/view/989.

Javeed, M. U., Shafqat Maria Aslam, Hafiza Ayesha Sadiqa, Ali Raza, Muhammad Munawar Iqbal, & Misbah Akram. (2025). Phishing Website URL Detection Using a Hybrid Machine Learning Approach. Journal of Computing & Biomedical Informatics, 9(01). Retrieved from https://jcbi.org/index.php/Main/article/view/989.

Javeed, M., Aslam, S., Farhan, M., Aslam, M., & Khan, M. (2023). An Enhanced Predictive Model for Heart Disease Diagnoses Using Machine Learning Algorithms. Technical Journal, 28(04), 64-73. Retrieved from https://tj.uettaxila.edu.pk/index.php/technical-journal/article/view/1828.

K. Vasudeva and S. Chandran, “A comprehensive study on facial expression recognition techniques using convolutional neural network,” in *2020 International Conference on Communication and Signal Processing (ICCSP)*, 2020, pp. 1431–1436, doi: 10.1109/ICCSP48568.2020.9182108.

L. M. Darshan and K. B. Nagasundara, “A survey on disguise face recognition,” *J. Chin. Inst. Eng.*, vol. 47, no. 5, pp. 528–543, 2024, doi: 10.1080/02533839.2024.0000000.

L. Zhang, B. Verma, D. Tjondronegoro, and V. Chandran, “Facial expression analysis under partial occlusion: A survey,” *ACM Comput. Surv.*, vol. 51, no. 2, pp. 1–49, 2018, doi: 10.1145/3158230.

Li, J., Hu, L., & Basheer, M. F. (2024). Linking green perceived value and green brand loyalty: a mediated moderation analysis of green brand attachment, green self-image congruity, and green conspicuous consumption. Environment, Development and Sustainability, 26(10), 25569-25587.

M. Garg and R. S. Prasad, Eds., *Affective computing for social good: Enhancing well-being, empathy, and equity*. Springer Nature, 2024.

M. R. King, “Prop'eau sable,” *Recherche-action en vue de la préparation et de la mise en œuvre du plan d'action de la zone des sables bruxelliens en application de la directive européenne CEE/91/676 (nitrates)*, 2024.

M.U. Javeed, M. S. Ali, A. Iqbal, M. Azhar, S. M. Aslam and I. Shabbir, "Transforming Heart Disease Detection with BERT: Novel Architectures and Fine-Tuning Techniques," 2024 International Conference on Frontiers of Information Technology (FIT), Islamabad, Pakistan, 2024, pp. 1-6, doi: 10.1109/FIT63703.2024.10838424.

Mahrukh Jaffar, “ONTOLOGY-BASED SENTIMENT ANALYSIS FOR REAL-TIME PRODUCT REPUTATION MODELING”, SES, vol. 3, no. 7, pp. 648–667, Jul. 2025.

Muhammad Usman Javeed, Hafiza Ayesha Sadiqa, Mahrukh Jaffar, Shafqat Maria Aslam, Muhammad Khadim Hussain, Zeeshan Raza, & Muhammad Azhar. (2025). A DEEP LEARNING APPROACH FOR SECURING IOT SYSTEMS WITH CNN-BASED PREDICTION OF WORST-CASE RESPONSE TIME. Spectrum of Engineering Sciences, 3(7), 376–385. Retrieved from https://www.sesjournal.com/index.php/1/article/view/599

N. Khan, A. Singh, and R. Agrawal, “Enhancing feature extraction technique through spatial deep learning model for facial emotion detection,” *Ann. Emerg. Technol. Comput. (AETiC)*, vol. 7, no. 2, pp. 9–22, 2023, doi: 10.33166/AETiC.2023.02.002.

Q. Q. Oh, C. K. Seow, M. Yusuff, S. Pranata, and Q. Cao, “The impact of face mask and emotion on automatic speech recognition (ASR) and speech emotion recognition (SER),” in *2023 8th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, 2023, pp. 523–531, doi: 10.1109/ICCCBDA.2023.10105092.

Qadir, F., Basheer, M. F., & Chaudhry, S. (2024). Transgender Entrepreneurs are Paving the Path of Social Entrepreneurship: Exploring Motivators of Entrepreneurial Intent. Journal of Business and Management Research, 3(3), 785-812.

R. M. Al-Eidan, H. Al-Khalifa, and A. Al-Salman, "Deep-learning-based models for pain recognition: A systematic review," *Appl. Sci.*, vol. 10, no. 17, p. 5984, 2020, doi: 10.3390/app10175984.

Shakeel, H. ., Akram, M. ., Javeed, M. U., Azhar, M. ., Aslam, S. M. ., Saifullah, & Mumtaz, M. T. . (2025). LncRNAs Disease: A text mining Approach to Find the role of lncRNA in Aging. Journal of Computing & Biomedical Informatics, 9(01). Retrieved from https://www.jcbi.org/index.php/Main/article/view/1000

T. Wehrle, S. Kaiser, S. Schmidt, and K. R. Scherer, "Studying the dynamics of emotional expression using synthesized facial muscle movements," *J. Pers. Soc. Psychol.*, vol. 78, no. 1, pp. 105–118, 2000, doi: 10.1037/0022-3514.78.1.105.

V. S. Amal, S. Suresh, and G. Deepa, "Real-time emotion recognition from facial expressions using convolutional neural network with Fer2013 dataset," in *Ubiquitous Intelligent Systems: Proceedings of ICUIS 2021*, Springer Singapore, 2022, pp. 541–551.

W. Wu *et al.*, "Look at boundary: A boundary-aware face alignment algorithm," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2129–2138.

Y. X. Tan *et al.*, "Recent advances in text-to-image synthesis: Approaches, datasets and future research prospects," *IEEE Access*, 2023, doi: 10.1109/ACCESS.2023.3298829.

Yin, X., Khan, A. J., Basheer, M. F., Iqbal, J., & Hameed, W. U. (2025). Green human resource management: a need of time and a sustainable solution for organizations and environment. Environment, Development and Sustainability, 27(1), 1379-1400.