International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

A Hybrid Stacked CNN Model with Weighted Average Ensembling for Effective Face Recognition

[1]Rishi Sharma, [2]Anjali Potnis
[1]Electronics and Communication, OIST, Bhopal, India
rishisturdy@gmail.com
[2]Electrical & Electronics, NITTTR, Bhopal, India
apotnis@nitttrbpl.ac.in

Abstract—The discipline of computer vision has given a lot of attention to facial recognition. Automated face recognition is extensively employed in various practical scenarios, including systems for streamlining immigration checkpoints, intelligent monitoring of visual data, and authentication of personal identity. Depending on the situation, it may be divided into the two separate duties of facial verification and face identification. This study proposed a hybrid stacked CNN model for face recognition system. The models have mastered the art of making their own inferences. The models are then combined to predict a class value using a cutting-edge method called weighted average ensembling. A more accurate estimate should be produced by the new assembly process. Pre-trained CNN models are used in our proposed method: AlexNet, Resnet50V2, and VGG-19, Yolov5, VGG-16 and ResNet50. When applied to Tufts dataset images, our suggested model successfully achieved 98.05% accuracy. We have also used the Discrete Wavelet Transform (DWT) method for denoising, SegNet for Image segmentation for better performance of the model proposed.

Keywords-Face Recognition; Deep Learning; Transfer Learning; Discrete Wavelet Transform

I. INTRODUCTION

Facial recognition has received a lot of attention in the field of computer vision. Several real-world applications rely heavily on automatic facial recognition, including verification, identification, automated clearance systems at immigration checkpoints, and intelligent visual surveillance. Depending on the context in which it is employed, facial recognition can be divided into two separate tasks; facial verification and face identification. Due to changes in facial expressions, positions, and lighting, face identification is an extremely difficult study subject in pattern detection [1] and computer vision. Face recognition systems must be developed by the industry to be effective and automated for a number of new applications, including commercial and law enforcement duties. Although many academics have spent years tackling the issue of facial recognition[2, 3, 4], there are still a number of problems that need to be resolved.

Since it's hard to identify human faces while they're conveying different emotions. The facial recognition model may be executed using a wide variety of methods. In this paper we have used a Tufts dataset that has great variety of face image in different modalities like computer sketch, 3D images, near- infrared imaging (NIR) images etc. Then Local Binary Patterns Histograms (LBPH) method is utilised for feature extraction. AlexNet, VGG-16, VGG-19, YOLOv5 convolutional neural networks are employed for comparison with the pretrained networks and the XGBoost technique is used for image classification in our face recognition system. ResNet-50 encoder- decoder is also used for improved face detection process. The ResNet-50 outperforms for encoding other image recognition algorithms[5] in terms of accuracy. The Haar cascade method is also utilised for facial recognition. The paper decided to adopt the Haar cascade classifier due to its impressive accuracy in detection, fast processing speed, and minimal occurrenceof false positives. A large number of both positive (facial)and negative (non-face) pictures were used throughout its training. Following the face identification procedure, XGBoost algorithm is used to compare and categorise the discovered face photos against the database images.

A. Author Contribution

- Tufts dataset has been used and has seven different image modalities that helped us to improve our face recognition system with maximum possible image format.
- De-noising of natural images by discrete wavelet transform (DWT) techniques. It is a commonly used technique for image denoising. It operates by decomposing an image into different frequency components using wavelet filters and then selectively modifying or removing the noise in the appropriate frequency bands.

TheHaar cascade classifier under goes training for feature extraction and is subsequently employed for face identification. This machine learning approach requires many examples of both good and bad pictures to properly train the classifier. Classifiers are chosen due to their

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

ability to achieve high detection accuracy, fast processing speed, and a low occurrence of false positives. We have also used predefined tools from OpenCV for classification.

- Different CNN intelligent paradigm like AlexNet, Yolov5, ResNet, VGG-16, VGG-19 and ResNet-50 are employed to automatically select features from the frequency bands extracted using DWT. ResNet-50 encoder-decoder has also been utilized to generate image-encoded vector.
- The fundamental steps in classification modeling involve acquiring a dataset, extracting features from independent variables, and utilizing them to predict a dependent variable or target class. The Extreme Boost (XGBoost) Classifier is used in this paper for further improvement in face detection and recognition process.
- To evaluate our proposed system, we employ the Tufts dataset, which is known for its challenging benchmarks. In contrast to other anomaly detection methods, our approach yields state of the art results. By leveraging DWT- based transfer learning and a feature extraction model strategy, we achieve an impressive accuracy of 98.05%.

B. Article Organization

The remaining sections of the paper are structured as follows: Section 2 provides an examination of existing methodologies and their related research. Section 3 outlines the identification of problems in face recognition systems and introduces our proposed model for improvement. Section 4 presents the basic system architecture of a face recognition system, detailing each step of its operation. This includes data description, face detection methods, and face recognition methods. Section 5 delves into a description of the proposed methodology framework. Section 6 evaluates the results obtained, while Section 7 thoroughly discusses the experimental results and provides a comprehensive comparison with existing techniques. Finally, in Section 8, we present the conclusion of our research work.

II. RELATED WORK

In section provide a quick summary of research done on facial recognition in the past few years. As computing power has improved, a number of researchers in the field of face identification and recognition have made significant advancements; several pieces of study have been conducted to investigate these developments.

Yang et al. [6] provided SR-CNN, a deep-face-recognition model that integrates many characteristics in order to recognise a variety of activities. The model they suggest is a fusion of in- variant texture feature (RITF). Wholesale retaining the benefits of the basic SIFT and RITF models, this approach enhances the twirl invariance of the conventional SIFT model. The authors employed the CNN model that includes a batch normalisation layer, the parametric rectified linear unit (PReLU) as the activation function, and cross-entropy cost function adjustments to the network parameters and tested over two self-generated dataset.

Hansen et al.[7] discovered the passive Radio Frequency Identification (RFID) system, which is used to identify pigs and other livestock at a farm. To combat the problem, they attempted to use the facial recognition method with pre-trained CNN models. The authors described the outcomes of three face recognition techniques used on a dataset of pig faces that were photographed in their natural habitat on a farm. They employed the pre-trained VGG facial recognition model as well as their own CNN, which was trained using data they had collected from a common web camera.

Jiangetal.[8]employed the widely used face detection dataset and benchmark (FDDB), IJB-A benchmark, and WIDER face data set as three well-known face detection benchmarks. The authors fine-tuned the pre-trainedVGG16modelin order toutilise these datasets in a quicker R-CNN model. The authors specifically used 10,000 iterations to optimise the face identification model on the training photos from each split of the IJB-A. The basic learning rate is 0.001 for the first 5,000 iterations and drops to 0.0001 during the last 5,000 iterations. The author performed 10,000 rounds of fine-tuning to match the IJB-A annotation styles to the regression branches of the Faster R-CNN model trained on WIDER.

Farfade et al.[9] presented a Deep Dense Face Detector (DDFD),a technique to recognize faces in a variety of orientations using a single deep convolutional neural network model without the need for pose/landmark annotation. To fine-tune AlexNet for face identification, the authors extracted training samples from the AFLW dataset, which comprises 21K photos with 24K face annotations. The authors used a sliding window strategy for their work since it is less difficult and independent of other modules like selective search. On the datasets such as PASCAL Face, Annotated Faces in the Wild (AFW), and FDDB, they further evaluated their face detection methodology. Their findings stated that, even without posture annotation or data on face landmarks, their results were better than those of R-CNN.

III. PROBLEM FORMULATION AND MOTIVATION

There are several challenges associated with face recognition systems such as illumination, occlusion, pose variation, and various strange facial expressions apart from that the major issues in the face detection system are noisy images and image segmentation. It is believed that a picture speaks a thousand words. One of the significant challenges in the field of image processing is the presence of noise that corrupts the image. During the processes of picture capture, compres- sion, and transmission, noise may enter the image. There are several other reasons, including hardware issues with the camera lens, inadequate processing power, etc. It is important to keep picture noise as low as possible so that the face recognition system can have a better image to process with greater

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

accuracy. For localised features like edges and curves, DiscreteWaveletTransformalgorithmprovidethebestresults.Depending on the frequency properties of the signal, wavelet transform divides the picture into many components [10].

IV. BASIC ARCHITECTURE OF FACE RECOGNITION SYSTEM

Our methodology involved in the face recognition system consists of some basic steps, including defining dataset, image pre-processing, face detection, feature extraction, classification, and face recognition. Figure **1**illustrates the proposed facial recognition system along with the stages involved in each task. Pre-processing of images speeds up the matching process and decreases processing time. Pre-processing is performed on facial photos so that they may be used for feature extraction. It is important to carefully evaluate the ideal picture size since various sizes provide different information. Image scaling is done to get a smaller data size, which in turn reduces processing time. Scaling the data values to a predetermined range (either -1.0 to 1.0 or 0.0 to 1.0) is a crucial part of the data preparation phase.

A. Data Description

The widespread use of many sensors in commonplace applications has given rise to the study of cross-modality face recognition. The assessment and training for data-hungry algorithms for machine learning are crucial to the development of face recognition systems. The Tufts Face Database contains a collection of over 10,000 photographs, consisting of 74 females and 38 males from more than 15 nations. The data set encompasses a wide age range, spanning from 4 to 70 years old. Notably, this dataset encompasses seven distinct image modalities, including visible, near-infrared, thermal, computerized sketch, LYTRO, recorded video, and 3D images. Table I provides a description of the Tufts dataset images, along with the methods used to prepare those categories of images. The Tufts Face Database's image collection took place in a 9x10ftphotography studio. In order to create a3D model of the human face, many 2D photos were taken of humans in controlled lighting circumstances with a variety of facial expressions and props. Computer-generated drawings were created using the FACEs sketch program. The 3D face image capture and reconstruction were aided by deriving nine camera locations. In addition, thermal facial photos were captured using specialised thermal and NIR camera systems, in addition to carefully calibrated temperature and illumination conditions. Figure **1** Images taken from the Tufts Face Data base that serve as examples of the comprehensiveness of the dataset. The demographics of the Tufts Database span the globe and span the generations and cultures of humanity. To gather and release the Tufts Face Database, a procedure approved by the institution's research board (IRB) was created. Information on the study's goals and methods, the database's public and private accessibility, and the safeguarding of participants' identities were all included in the IRB protocol.

B. Face Recognition

The Face Recognition model identifies a picture of a face by comparing it to pictures of faces already in a database. The process of facial recognition may be broken down into three distinct phases: face detection, feature extraction, and facial recognition. Faces in an image may be located with the help of a face detection model.

There are certain challenges in recognising the same faces in various orientations, which is why the utilization of face alignment techniques is implemented to enhance the effectiveness of face recognition models, especially within specific libraries. After a face has been aligned, feature extraction may be performed to retrieve the most important characteristicsforrecognition.Detectedareasofthefaceareconverted into vector points using the appropriate model. Finally, face Recognition is accomplished by using a classification model based on a confidence score to compare observed face-to-face characteristics with those already recorded in a database. Figure 1 depicts the whole process followed by the Face Recognition system.

C. Face Detection

Many different kinds of face detection algorithms are putto use in fields as diverse as security monitoring, gaming, and human-computer interaction. The camera's face recognition feature can pick out people's faces in still images and movies and identify them from other things. In earlier work[11], the authors created an object identification technique usingthe Haar cascade classifier. The Haar cascade classifier undergoes training for the purpose of extracting features, which are subsequently utilized for identifying faces. This machine learning approach requires a large number of examples ofboth good and bad pictures to properly train the classifier. Classifiers are employed due to their ability to achieve high accuracy in detection, operate at fast speeds, and maintain a low rate of false positives. One such approach[12] uses the following equation to get the image value of each individual basic feature:

$$P = \text{black} - \text{white} = \frac{1}{n}\sum_{\text{black}}^{n} P(x) - \frac{1}{n}\sum_{\text{white}}^{n} P(x), \quad (1)$$

Where n represents the pixel value and P(x) is the obtained value from given input image, which is shown in the Figure 1. Numerous detection algorithms have been created, each with its own unique processing requirements and approach to detection. The Haar classifier is an effective tool for improving rejection cascade designs. For the first time, this architecture was put to use in a high-accuracy, real-time face detector [13]. The Haar characteristics are very useful for face recognition because of their ability to pick up on sharp angles and lines. Networks and face detection algorithms that rely on edge and line characteristics can only identify objects with sharp corners or sharp edges.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

TABLE I:Tufts data set description: Image types and method used to prepare face data

Category	File format	Equipment	Methods	Average file size
Computer Sketches	.jpg	FACES 4.0Software	Researchers use the programme to choose a group of prospectiveface components from the database based on their own observation or recollection	76 KB
3D Images	.ply	Quad camera	Open-source structure-from-motion methods were used to recreate the 3D models	21 MB
Thermal	.jpg	FLIR Vue Pro camera	There were five different poses requested of each participant: blank face, grin, eyes closed, extreme shock, and sunglasses	187 KB
NIR	jpg	Night vision quad camera	An 850nm Infrared 96 LED light setup was used to keep the brightness consistent for NIR imaging.	13 MB
LYTRO	.lfr	LYTRO ILLUM 40 Megaray Light Field camera	Approximately 9 evenly spaced cameras were positioned in a semicircle around the subject	53 MB
Video	.h264	Visible fieldcamera	The camera was panned in a semicircle around the subject.	40 MB

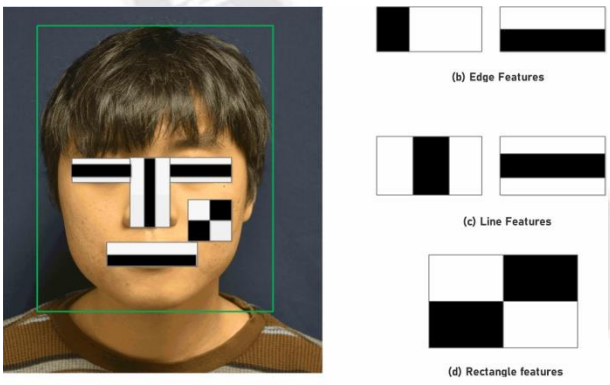

Fig.1:FeaturesExtraction

V. PROPOSED METHODOLOGY

To illustrate the efficacy of our suggested face recognition technique, we have used a well-known dataset that is Tufts. This dataset is described in the above part of the paper. Our proposed model comprises basic steps which involve data preprocessing, image encoding, data splitting, feature extraction, classification, and face recognition. Figure 2 presents the visual representation of the block diagram that showcases the structure and components of our proposed model.

The suggested approach for face recognition starts with the initial phase of getting face photos or images with low quality and then denoising the picked image. In this paper, we have used DWT (Discrete Wavelet Transform) technique for denoising the available images. The removal or preservation of picture noise significantly influences subsequent digital image processing tasks, including image segmentation, edge detection, feature extraction, and more. The quality of noise removal directly impacts the accuracy and effectiveness of these subsequent processing steps. We have used nonlinear denoising techniques that rely mostly on thresholding the DWT coefficients, which have been impacted by additive white Gaussian noise.

$$Uc = \sigma\sqrt{(2\log L)} \qquad (2)$$

Where Uc is the universal threshold value, L is the length of the data, and σ is the expected noise variance of the data based on the equation. Due to the fact that it does not

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

statistically check each piece of data, universal thresholding is non-data dependent. However, given that its formulation includes parameters like L and, it is unquestionably an adaptive threshold approach. Each level of breakdown consists of two phases. Phase I filters horizontally and phase 2 filters vertically as shown in Figure 3. When an input image of size NXN undergoes first-level decomposition, it yields three subbands: LH, HL, and HH. Each subband has dimensions of N/2XN/2. The LL band serves as the input for the secondlevel decomposition, while the three N/4XN/4 subbands LLLH, LLHL, and LLHH serve as the outputs. The LLLL band serves as the third-level decomposition's input, while the four N/8XN/8 subbands (LL)2LL, (LL)2LH, (LL)2HL, and (LL)2HH serve as its outputs thresholded [14, 15]. Inverse wavelet transform is used to rebuild the picture after thresholding.

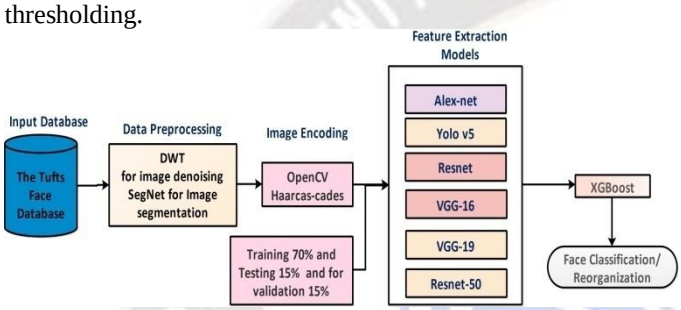

Fig.2:Block diagram of the proposed model for face recognition

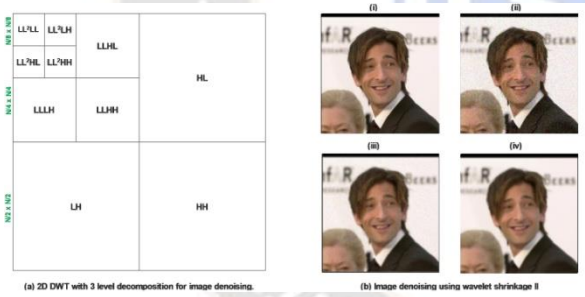

Fig.3:Pre-processing for image-denoising using DWT

In the next phase, we have worked on the image segmentation process and for this, we have used SegNet network. SegNet is a deep neural network with a three-depth encoder decoder. The convolution layers of the VGG16 network are similar to the encoder of SegNet. In SegNet the segmentation mask is created by the decoder using pooling indices from the maximum pooling of the matching encoder. To decrease complexity, developers eliminated the completely linked layers, which brought the encoder sector's parameter count down from 1.34e+8 to 1.47e+7 [16]. Any random multichannel images or feature map, such as RGB, RGBD, a map of normals, a depth map, etc., may be used as the SegNet's input. As a first step to the input processing, local contrast normalisation (LCN) is carried out. This action has various benefits, it reduces the dynamic range by correcting for uneven scene's lighting. The network learns category form by emphasising edges, and convergence is improved by de-correlating the input dimensions. LCN is carried out separately for each modality. For training we have used Limited-Memory Broyden-FletcherGoldfarb-Shanno (L-BFGS) algorithm, it provides a quicker and more reliable convergence. Additionally, it works well in big batches, which is advantageous for maximising the throughput of potent GPUs.

A. Pre-trained CNN

Image categorization, within the realm of computer vision, holds significant importance as a fundamental and practical field of research. Because of its quick and precise feature extraction function and fully trainable network structure, convolution neural networks (CNNs) have seen tremendous success in the area of image classification in recent years. It was in the 1990s when the foundations for what would become CNN were laid.

B. Alexnet

In order to pave the way for contemporary CNN, Lecun et al. [17]created LeNet-5. In 2012, Krizhevsky et al. [18] developed an expanded and improved version of LeNet-5 called AlexNet, which went on to win the ImageNet Competition and spark a spike of interest in CNN research.

C. VGGNet

Oxford University proposed the VGGNet in 2014, using smaller 3 x 3 size convolution kernels and stacking them to replace larger size convolution kernels. This allows for improved network performance with fewer trained parameters. In the ImageNet Stereo Vision Challenge (ILSVRC) localization and classification competitions, the VGGNet placed first and second, respectively.

D. ResNet

In 2015, by He et al. [19], ResNet was able to address gradient disappearance in deep CNN, hence enhancing its ability to correctly classify images. Figure 4 depicts the basic architecture of ResNet model. The layers of CNN are modified by different authors as per their experimental requirement and for the improvement of result.

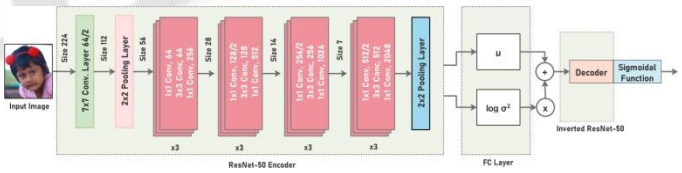

Fig.4:TypicalResNet-50encoder-decoderArchitecture

E. Yolov5

YOLO-5 (You Only Look Once) is another model that we have used for real-time object detection. YOLO is cuttingedge deep learning framework for object detection for realtime

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

object identification [20]. A variety of things should be able to be detected using quick, precise object detection. The YOLO frameworks for detection are becoming faster and more precise with the aid of neural networks. The fundamental concept behind YOLO is to feed the network the complete image as input and have it immediately return the location of the bounding box and its corresponding category [21]. YOLO predicts each bounding box by considering the features of the entire image. Each bounding box is associated with five predictions and confidences, which are related to the grid unit located at the center of the bounding box. Based on the anticipated width and height of the complete picture, they make up the fundamental frame of YOLO.

F. LBPH Algorithm

The OpenCV library provides several feature extraction and recognition methods that may be used for free. Therefore, eigenfaces, fisher-faces, and local binary pattern histogram (LBPH) [22] are the most used approaches. The LBPH algorithm is superior than the other two because it is more adaptable because of its adaptability in recognizing both frontal and side faces. Although LBPH has a low computing complexity, it is nonetheless susceptible to flaws in areas like rotational fluctuations, facial analysis and identification, and gray-scale variability [23]. These benefits make LBPH applicable to systems with limited computational resources. Figure 5 illustrates the method by which a picture is converted into a decimal number inside the LBPH algorithm used in its calculation. The threshold is the value of the central pixel, which may be calculated by first converting the pixel value to binary and then to decimal. A pixel's binary value is 1 if its value is at or over the threshold; otherwise, it is 0. The algorithm then generates a threshold result in the form of a linear binary pattern (LBP), albeit this value may be transformed into a histogram by multiplying it by two binary numbers and a decimal number. A numeric value between 0 and 255 is shown in each cell of the histogram. The output from each individual histogram is then merged to form the final histogram.

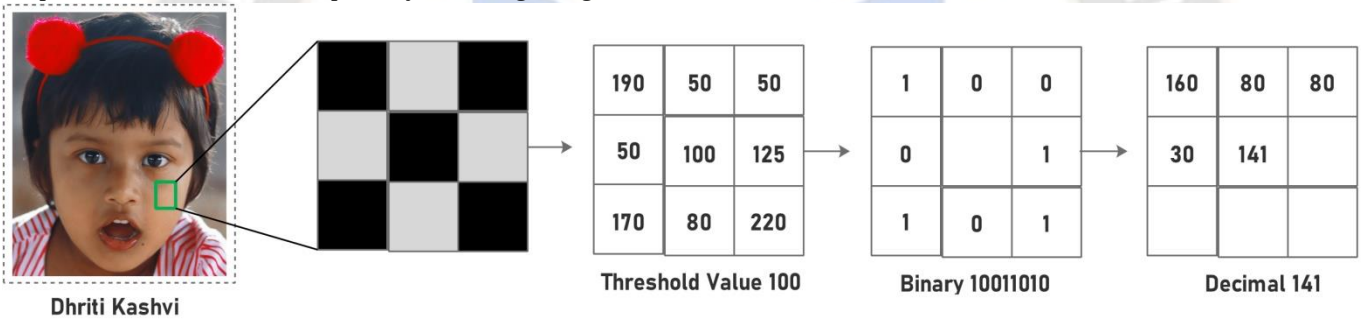

Fig. 5: Grayscale to Decimal Conversion

VI. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of our proposed method, we conducted a series of experiments using the Tufts datasets mentioned earlier in the preceding section. The dataset is extensive and comprises both conventional thermal and polarimetric thermal modalities. This necessitates two separate cross-modal face verification tasks: conventional thermal to visible and polarimetric thermal to visible. However, due to the absence of polarimetric thermal images in the Thermal Paired Face Database of the Tufts Face Database, our study focuses solely on thermal-to-visible cross-domain face verification using these datasets. We assess and contrast the suggested method's performance with the following current cutting-edge techniques [24, 25, 26, 27]. Also the multi-spectral databases that were used in Tufts dataset and the tests that were run to evaluate the effectiveness of the proposed method are both included in this section. The presentation attack detector is introduced and assessed in a preliminary test. The ResNet CNN layers that should be modified to raise the calibre of the face embedding are then the subject of our research. Then, using the 256-d embedding, we examine which classifiers are better at classifying a person's identification. The most advanced techniques for multi-spectral face recognition are compared to the best classifier in the final analysis. We have then applied skin detector which works on pixel level and the normalised difference between all face channels is calculated in the this stage as Sd[da, db] $Sd[d_a,d_b]$.

$$Sd\,[d_a, d_b] = \left(\frac{d_a - d_b}{d_a + d_b}\right) \tag{3}$$

where the number of channels id denoted by n accessible in the presentation attack detection module and d is the pixel intensity value for channels a and b with $a \leq b \leq n$. The normalised differences fall within the range 1 Sd[da, db] + 1. The skin detector categorises the pixels as "skin" or "not skin" based on the normalised difference values. For pictures in the VIS, NIR and LWIR spectral bands in Tufts dataset, the following values were used to determine the range of values to identifypixels as skin or not-skin: (76, 51, 65) (Sd[d1, d2], Sd[d1, d3], Sd[d2, d3]) (131, 140, 127). The skin pixels are represented by a binary map in which 1 denotes skin and 0 denotes not-skin. The amount of facial landmarks that the presentation attack detector considers to be skin is calculated with the help of this binary map.

The outcome of pre-processing stage serve as the foundation of the face recognition modules. A facial pictures

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

are normalized, resized, and cropped to fit the proposed CNN's dimensions. The picture is then analyzed by the CNN to obtain its 256-d embedding vectors. The identity of the person visible in the images then be determined by combining the embedding vector with a classifier. A number of experiments were conducted in order to determine which layer should be fine-tuned using domain-specific data and which kind of parameter of classifiers provide the highest performance. In the denoising process we have used Adam optimization algorithm.

The performance of the proposed model is analyzed using various evaluation metrics, including accuracy, precision, F1score, specificity, sensitivity, and positive likelihood ratio (+LR). The formulation of these metrics and their corresponding parameters are presented in Table II. The parameters for the confusion matrix are detailed in Table III.Analysis based on Receiver Operating Characteristic (ROC) Curve

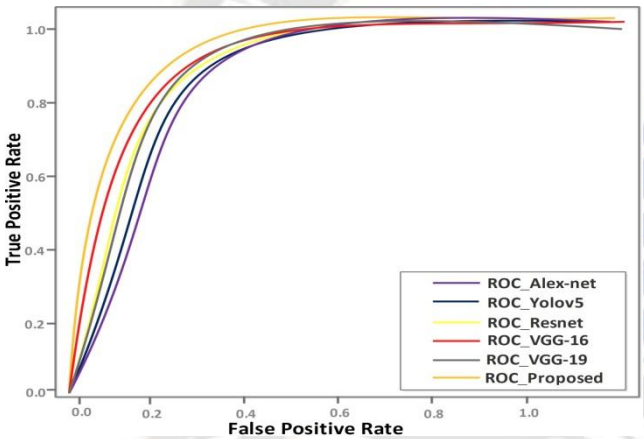

Fig.6:ROC of Model vs Proposed Model

The confusion matrix is employed to evaluate the receiver operating characteristic (ROC) curve and determine the success of identification. The ROC curve calculates the proportion oftrue and false positives in the test data, with the true positive rate on the Y-axis and the false positive rate on the X-axis. Figure 6 illustrates the ROC curve for the proposed model in comparison to other models. The area under the ROC curve (AUC) is a two-dimensional measure of the model's performance relative to alternative models, indicating the degree to which the proposed model outperforms them.

TABLEII:FormulaeofEvaluationMetrics

Parameter	Formula
Positve Predicated value (PPV) or Precision	$\frac{TP}{TP+FP} * 100$
Recall or Sensitivity	$\frac{TP}{TP+FN} * 100$
Specificity	$\frac{TN}{FP+TN} * 100$
F1-Score	$\frac{2*(PPV*Sensitivity)}{PPV+Sensitivity} * 100$
Accuracy	$\frac{TP+TN}{TP+FP+FN+TN} * 100$
Positive Likelihood Ratio(+LR)	$\frac{Sensitivity}{100-Specificity}$
FPR	$\frac{FP}{FP+TN} * 100$
FNR	$\frac{FN}{FN+TN} * 100$
FDR	$\frac{FP}{FP+TP} * 100$

B. Analysis based on Convergence Curve

A convergence curve illustrates the optimal value of the learning parameter by showcasing the relationship between the model's accuracy and the loss function during the training and testing phases. Figure 7 depicts the convergence curves of the proposed and alternative models. The sub-figures represent 7a AlexNet, 7b Yolo5, 7c VGG-16, 7d VGG-19, 7e ResNet and 7f proposed approach respectively. The 200-epoch training and validation accuracy curves for proposed models are displayed. The training and validation accuracy of the proposed model exhibit a faster convergence rate and achieve an accuracy of 98.05%. This accuracy surpasses the accuracy achieved by previously generated models.

TABLE III : Performance Evaluation:Model comparison metrics for Tufts dataset

Models	**Accuracy**	**Precision**	Sensitivity	**F1Score**	**Specificity**	**+LR**	**FPR**	**FNR**	**FDR**
AlexNet	85.15	89.70	89.83	89.80	89.80	8	11.02	08.12	34.61
Yolov5	94.60	95.90	95.64	95.90	94.90	19	04.64	03.38	16.40
ResNet	95.20	96.09	95.77	95.50	94.70	18	03.47	02.70	11.79
VGG-16	97.10	96.53	97.74	97.14	96.45	28	01.61	02.12	05.89
VGG-19	97.54	97.32	97.80	97.56	97.27	35	01.28	01.87	04.64
Proposed	**98.05**	**98.31**	**98.17**	**98.15**	**97.15**	**43**	**00.65**	**01.89**	**02.21**

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

VII Discussion

In this research, we use Adam optimizer to test the model with varying epochs and batch sizes until we get a satisfactory outcome, as seen in Table III. For this research, we employthe Adam optimizer to assess the model's performance by testing it with different combinations of epochs and batch sizes. We iterate this process until achieving a satisfactory outcome, as indicated in Table III. The AlexNet model obtains 85.15% accuracy, Yolov5 gets 94.60% accuracy, ResNet gets 95.20% accuracy, VGG-16 gets 97.10% accuracy, VGG-19 gets 97.54% and our proposed model gets 98.05% accuracyin this experiment. We have measured the efficacy of our approach using the roc and convergence curves. The ResNet network enhanced with XGBoost performs more effectively. All machine learning algorithms strive to achieve one thing: a lowerloss.Onesynonymforlossfunctionisthecostfunction.
When compared to VGG-16 (training loss = 0.024%) and LBPH (training loss = 0.045%), our suggested research finds that ResNet-50 provides the most consistent results. Facial recognition is simulated using Tufts dataset to identify and recognise individuals.

VIII Conclusion

Our suggested model for a facial recognition system makes use of the publicly accessible Tufts datasets, and it has been successfully applied in this research. Based on the photographs in its database, the face recognition system can distinguish between recognized and unfamiliar faces. In our research, we employed multiple techniques for face detection and feature extraction. For classification, we utilized pre-trained CNN models such as ResNet-50 and VGG-16 along with an XGBoost classifier. Additionally, we incorporated the LBPH algorithm for face recognition, which presents opportunities for further improvement. Comparing our proposed model to the other two methods, we observed that it achieved the highest recognition accuracy of 98.05%.

In the future, we plan to make the proposed model more robust and reliable by further reducing the training and validation loss thereby increasing the validation accuracy. Additionally, more datasets can be collected from other sources to includeas many edge cases as possible to train the model.

References

[1] Clint Feher, Yuval Elovici, Robert Moskovitch, Lior Rokach, and Alon Schclar. User identity verification via mouse dynamics. Information Sciences, 201:19–36, 2012.

[2] Huaizu Jiang and Erik Learned-Miller. Face detection with the faster r-cnn. In 2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017), pages 650–657. IEEE, 2017.

[3] Sachin Sudhakar Farfade, Mohammad J Saberian, and Li-Jia Li. Multi-view face detection using deep convolu- tional neural networks. In Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, pages 643–650, 2015.

[4] Maxim Kimlyk and Sergei Umnyashkin. Image denois- ing using discrete wavelet transform and edge infor- mation. In 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), pages 1823–1825. IEEE, 2018.

[5] Paul Viola and Michael J Jones. Robust real-time face detection. International journal of computer vision, 57:137–154, 2004.

[6] Soe Sandar and Saw Aung Nyein Oo. Development of a secured door lock system based on face recognition using raspberry pi and gsm module. Development, 3(5), 2019.

[7] Mohammad J Saberian and Nuno Vasconcelos. Boosting classifier cascades. In NIPS, volume 23, pages 2047– 2055, 2010.

[8] S Grace Chang, Bin Yu, and Martin Vetterli. Adaptive wavelet thresholding for image denoising and compres- sion. IEEE transactions on image processing, 9(9):1532– 1546, 2000.

[9] S Kother Mohideen, S Arumuga Perumal, and M Mo- hamed Sathik. Image de-noising using discrete wavelet transform. International Journal of Computer Science and Network Security, 8(1):213–216, 2008.

[10] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transac- tions on pattern analysis and machine intelligence, 39(12):2481–2495, 2017.

[11] Yann LeCun, Bernhard Boser, John Denker, Donnie Hen- derson, Richard Howard, Wayne Hubbard, and Lawrence Jackel. Handwritten digit recognition with a back-propagation network. Advances in neural information processing systems, 2, 1989.

[12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6):84–90, 2017.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[14] Christian Szegedy, Alexander Toshev, and Dumitru Er- han. Deep neural networks for object detection. Advances in neural information processing systems, 26, 2013.

[15] Jia Yao, Jiaming Qi, Jie Zhang, Hongmin Shao, Jia Yang, and Xin Li. A real-time detection algorithm for kiwifruit defects based on yolov5. Electronics, 10(14):1711, 2021.

[16] Raman Sharma, Dhanajay Kumar, Vaishali Puranik, Kri- tika Gautham, et al. Performance analysis of human face recognition techniques. In 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), pages 1–4. IEEE, 2019.

[17] Song Ke-Chen, YAN Yun-Hui, CHEN Wen-Hui, and Xu Zhang. Research and perspective on local binary pattern. Acta Automatica Sinica, 39(6):730–744, 2013.

[18] Naser Damer, Fadi Boutros, Khawla Mallat, Flo- rian Kirchbuchner, Jean-Luc Dugelay, and Arjan Kui- jper. Cascaded generation of high-quality color visible face images from thermal captures. arXiv preprint arXiv:1910.09524, 2019.

[19] Aravindh Mahendran and Andrea Vedaldi. Understand- ing deep image representations by inverting them. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 5188–5196, 2015.

[20] Khawla Mallat, Naser Damer, Fadi Boutros, Arjan Kui- jper, and Jean-Luc Dugelay. Cross-spectrum thermal to visible face recognition based on cascaded image syn- thesis. In 2019 International Conference on Biometrics (ICB), pages 1–8. IEEE, 2019.

[21] Benjamin S Riggan, Nathaniel J Short, and Shuowen Hu. Thermal to visible synthesis of face images using multiple regions. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pages 30–38. IEEE, 2018.