Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 22s (2025)
https://www.jneonatalsurg.com

Criminal Facial Recognition Based on Multi Stage Progressive V-Net and MTCNN with NASnet Architecture

S. S. Beulah Benslet[1*], P. Parameswari[2]

1Research Scholar, Department of CS, Karuppannan Mariappan College, Muthoor, Chettiyarpalayam

Email ID: beulahabraham15@gmail.com

[2]Principal, Palanisamy College of Arts, Perundurai-Erode Road, Tamil Nadu 638052, India

Email ID: paramtech20@gmail.com

[*]**Corresponding Author:**

S. S. Beulah Benslet

Research Scholar, Department of CS, Karuppannan Mariappan College, Chettiyarpalayam, Muthur, Tirupur District, Tamil Nadu, 638105

Email ID: beulahabraham15@gmail.com

Cite this paper as: S. S. Beulah Benslet, P. Parameswari, (2025). Criminal Facial Recognition Based on Multi Stage Progressive V-Net and MTCNN with NASnet Architecture. *Journal of Neonatal Surgery*, 14 (22s), 621-637.

ABSTRACT

Criminal Face Recognition is the ability to detect and recognize a criminal by their facial characteristics. Most crucial duty of the police searching for the offenders is criminal identification. But since the police have to look for it everywhere, it is the hardest and time-consuming duty. Cities and other public areas with a high population density will provide greater challenges. Manual identification methods occasionally provide additional information about offenders. However, there is no requirement for monitoring when using an automatic identification system (AIS) in a public setting. Because the MIS is taking longer time, it cannot adequately focus on everyone. So, automatic identification is essential for recognizing the criminal face. A distinct machine learning method is the automated prediction model that is most frequently employed to identify criminal faces. However, attaining accurate prediction with lesser error probability is quite difficult using the existing network models. To overcome these issues, deep learning algorithm has been developed for identifying the criminals. Initially the criminal and non-criminal images are collected and pre-processed using a multi-stage progressive V-net approach. After that, the MT-CNN algorithm extracts features from the pre-processed image. The MTCNN (Multi-Task Cascaded Convolutional Networks) algorithm finds and recognizes faces in digital images or videos by employing a cascading sequence of convolutional neural networks (CNNs). Using the NASnet method, the segmented image is finally categorized to detect both criminal and non-criminal activity. The experimental analysis shows that the suggested strategy achieves 97.2% accuracy, 96.9% f1-score, and 2.60% FDR. Thus, the automated criminal detection system not only offers the police enormous convenience in identifying offenders, but it also saves them time.

Keyword: *Criminal Face Recognition; Artificial intelligence; criminal activity; multi-stage progressive V-net; MT-CNN*

1. INTRODUCTION

Recognition system is a system that can recognize a subject by its metrics. The recognition system is widely used for human authentication in many ways such as fingerprints, eye-based authentication by scanning the iris and retina, voice recognition, hand geometry and facial recognition. Among these recognition systems, the facial recognition system is targeted [1]. Facial recognition system is an architecture for identifying a person using AI technology and facial biometrics. Facial recognition is possible to identify individuals in images, films and in real time. By training the person's image with the AL technology, it can learn and identify if the person's image is spotted again. The AI technology learns the face by selecting the features of the person's face [2]. Most commonly used features are calculating the distance between two points and the depth of the elements, such as distance between two eyes, depth of the eyes, distance between the forehead and the shape of chin, eyes and lips. Calculating the distance between the feature points creates a template for each individual. Facial recognition systems are mainly used for authentication, identification and information technologies [3]. Through these technologies, the facial recognition system is targeted at identification technology. The facial recognition system based on subject identification works by training a set of images and compare the features of the trained image and tested image. The future of the facial recognition system depends on the balance of innovation and creativity in the technology [4].

Deep diving into the concept of facial recognition systems, the existing and proposed approaches are based on the classification of criminals and non-criminals. The identification of criminals and non-criminals can be done in two ways such as manual and automated. Police personnel scan people for identity under the Manual identity System (MIS) in public areas [5]. It takes a long time to provide the right care, and there's a danger that criminals won't get the treatment they need since they can easily flee after becoming aware of the police. However, there's no need for public surveillance when using an automated identification system (AIS). Before the classification of the images, there are various processes to be undergone by the image, such as pre-processing, feature selection, etc. [6]. Process of facial recognition starts with the collection of data about the person according to their classification output in the form of an image. Then, to remove the noise from the image and enhance the quality of the image, numerous pre-processing techniques are used. After pre-processing, the images are feature-extracted by many feature-extraction techniques [7]. Finally, the extracted features are classified by several machine learning algorithms and deep learning algorithms. Facial recognition systems are rapidly becoming more widespread and relevant [8].

Some of the existing approaches by which the classification of criminals and non-criminals are classified by convolutional neural network (CNN), Deep Neural Network (DNN), standard feed forward neural network (SNN), deep belief network (DBN), K-Nearest Neighbours (KNN) and Haar Cascade Classifier (HCC) [9]. Every current system categorizes the image with advantages and disadvantages, but improvements are still needed in the present systems' accuracy and efficiency. There are many methods that are already designed with low recognition capability, high false alarm rate [10]. In order to overcome the drawbacks of existing approaches, facial recognition system using machine learning, pre-processing, feature extraction and feature selection methods are used. The major contribution of the proposed system is given below.

- Deep learning algorithm is developed for automatic identification of criminal and non-criminal faces.
- Multi stage Progressive V-net algorithm (MPRNet) is used for pre-processing in order to enhance the quality of the original image for further processing.
- MT-CNN algorithm is used to extract the pre-processed image which generates the facial landmarks of the given images.
- NASnet architecture is used for predicting the criminal face based on the extracted facial features.
- False discovery rate, hit rate, selectivity etc. are calculated for showing the superiority of the proposed model.

The remaining section of this document includes: Section 2 shows the list of existing approaches for NASnet designed based on image data. The architecture of the proposed algorithm is discussed in Section 3. Section 4 shows the Result and graphical representation of the proposed system. Sections 5 and 6 comprise the conclusion and reference of the proposed approaches.

2. LITERATURE REVIEW

Many studies uses many techniques and classifies the criminal and non- criminal. A short form of many existing system is carried out below.

Hashemi and Hall [11] have developed a convolutional neural network and standard feed forward neural network for the purpose of detecting criminal tendencies using facial recognition technology. The data was pre-processed by a grayscale mugshot and then the images were classified by SNN and CNN. The dataset was taken from National Institute of Standards and Technology (NIST). The accuracy of the existing approach was 97% with CNN and 89% with SNN. Although there are certain drawbacks to the current method, such as the short dataset size and the fact that the photographs classified as criminal and non-criminal originate from various sources.

Sandhya at el. [12] have suggested a Deep Neural Network (DNN) for criminal detection with depends on facial recognition system. The approach employs an auto-encoder model, whose encoder portion is utilized to match the recorded facial images with a single-shot multibox detector, to detect faces. With training data, the current method obtains an accuracy of 92.50%, and with testing data, it reaches 90.90%. Despite this, the model lacks the flexibility to detect suspicious behaviour in addition to identifying offenders and does not employ sophisticated ensemble methodologies.

Sakiba at el. [13] have introduced a convolutional LSTM and YOLOv7 for detecting criminals dealing with facial recognition systems. The data image was pre-processed by normalization and data partitioning. Then the data was classified by convolutional LSTM and YOLOv7. 91% of accuracy for convolutional LSTM and 89.2% for YOLOv7 have been achieved by the current approach. But also, the approach doesn't use the aggregation of the two models with ensemble learning for classification.

Ahire at el. [14] have designed a support vector machine (SVM) for detecting criminals based on facial recognition systems. The Police Department's planned automatic criminal identification system aims to improve and modernize criminal distinguishing into a more practical and efficient method. The accuracy of the existing approach was 80% for SVM. Even though the approach contains many disadvantages, such as alarms for the criminal detection system.

Jahan at el. [15] have developed a K-Nearest Neighbours (KNN) for detecting criminals based on facial recognition systems. The current strategy is to take into account live streaming on surveillance systems in order to identify human faces from real-time video feeds and enhance security concerns in the campus region. 91.05% of accuracy for KNN in the existing approach. But also, the tensor flow of the proposed system is not perfect, so the accuracy of the system is lower.

You [16] have suggested a deep belief network (DBN) for detecting criminals based on facial recognition systems. For both feature extraction and classification, a face recognition method is created. Also, Deep Belief Network (DBN)'s Discrete Lion Swarm Optimization Algorithm (DLSA) is designed for facial identification. The accuracy of the existing approaches is 97.73% when using DBN. Despite the fact that the current method has several drawbacks, such as a tiny dataset and the fact that the photos of criminals and non-criminals originate from separate sources.

Patil et al. [17] have introduced a conventional neural network (CNN) for detecting criminals based on facial recognition systems. Artificial intelligence systems are being used as a result of technological advancements since they can recognize and comprehend emotions from facial features. The accuracy of 88% for CNN in the existing system. Furthermore, the method does not employ ensemble learning to aggregate the two models for classification.

Kulkarni et al. [18] have designed a Haar Cascade Classifier (HCC) for facial recognition systems for criminal detection. The existing approach has many pre-processing techniques, such as grayscale, image binarization, and image cropping. Following the feature extraction approach, HCC is used to classify the facial features. The facial recognition system's accuracy rate was approximately 87% after using the Haar Cascade Classifier. Even though the existing approach contains drawbacks such as over-fitting.

Several existing approaches developed to classify the criminal and non- criminal even though it contains many drawbacks. While the current method has certain drawbacks, like a short dataset and disparate sources for criminal and non-criminal photos [11]. While the model lacks flexibility in detecting suspicious activity in addition to identifying offenders, it does not require sophisticated ensemble techniques [12]. Furthermore, the method does not employ ensemble learning to aggregate the two models for classification [13]. Despite the fact that the strategy has numerous drawbacks, including alerts for the criminal detection system [14]. However, the suggested system's tensor flow isn't flawless, which reduces the system's accuracy [15]. Even if the existing approach has a number of shortcomings, including a small dataset and the fact that the images of criminals and non-criminals come from different sources [16]. Additionally, the technique does not combine the two models for classification using ensemble learning [17]. In addition, the technique does not use ensemble learning to combine the two models for classification [18].

PROPOSED METHODOLOGY

Facial recognition is used for identifying criminal and non- criminal in the proposed system. Some of the existing approaches by which the classification of criminals and non-criminals are classified by convolutional neural network (CNN), Deep Neural Network (DNN), standard feed forward neural network (SNN), deep belief network (DBN), K-Nearest Neighbours (KNN) and Haar Cascade Classifier (HCC). Numerous techniques now in use have poor detection ability and a high false alarm rate. A NASnet algorithm is created to address this problem. Architecture of the proposed criminal and non- criminal face identification is shown in Figure 1.

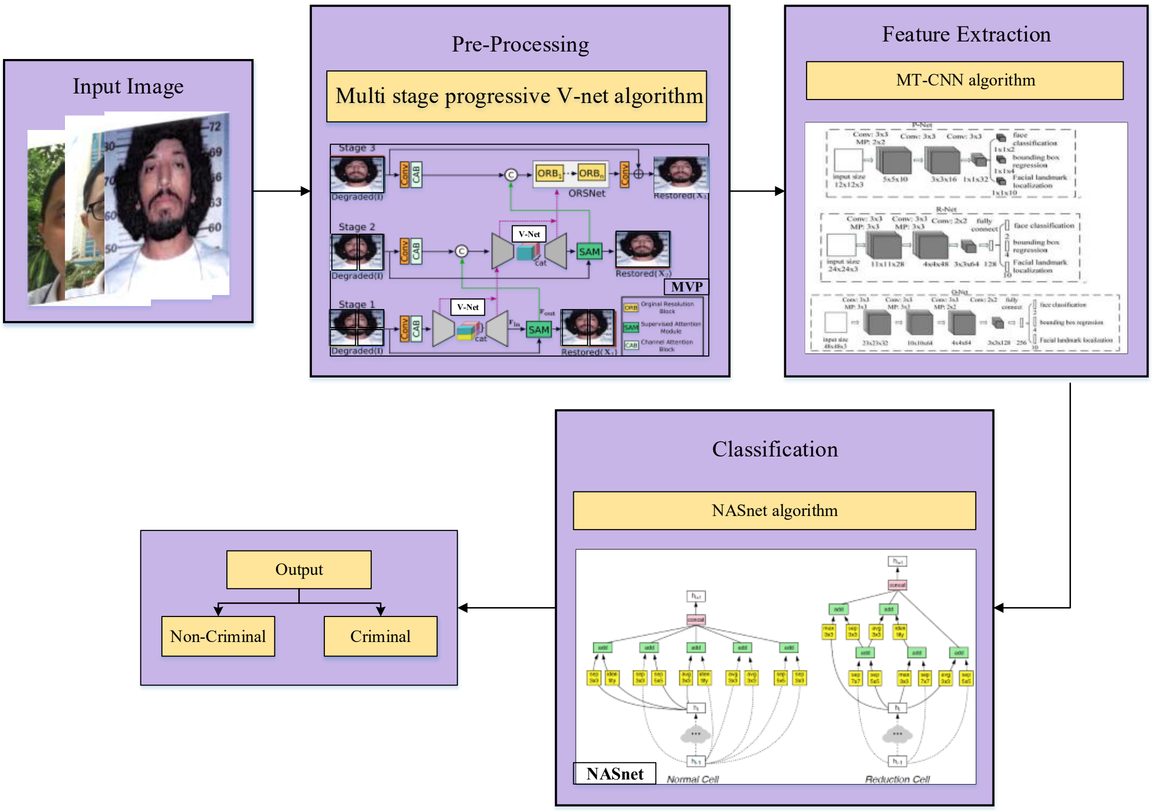

Figure 1. Architecture of proposed methodology.

Initially, the data is collected from photos and videos in the form of images. The proposed approach contains many steps for criminal identification. The collected data is initially pre-processed by a Multi stage progressive V-net algorithm. Then the pre-processed image is feature extracted by the MT-CNN algorithm. A cascading series of convolutional neural networks (CNNs) is used by the MTCNN (Multi-Task Cascaded Convolutional Networks) algorithm, a deep learning-based face identification and alignment technique, to find and recognize faces in digital photos or videos. Finally, the segmented image is classified by using the NASnet algorithm. The NASNet model is constructed by appending many convolutional layers, pooling layers, and activation functions to the picture. Up until the last layer produces a label, each layer's output is passed along as input to the layers around it. Thus, the system is effectively used for detecting criminal and non- criminal based on the NASnet algorithm.

Pre-Processing

Image pre-processing is the act of preparing and transforming unprocessed data into a format that can be used to train ML models. Preparing the data so that the neural network can process it simply is the first step in most deep learning workflows. Multi stage progressive V-net algorithm is used as an image pre-processing technique in the proposed system.

Multi stage progressive V-net algorithm

Three phases make up MPRV-Net's progressive picture restoration technique. Initial two are built using a conventional V-net-based encoder-decoder architecture, which gathers all related information from the original image [19]. Original Resolution Sub-Networks (ORSNet) is employed which operates at the original image resolution and producing spatially precise outcomes for the pixel-by-pixel correspondence among the input and output images.

Instead of only cascading numerous stages, explore a supervised attention module in between each step. The module rescales the characteristics maps while monitoring ground truth images, prior to moving them from one stage to another. Using the intermediate multi-scale contextualised characteristics of the preceding sub network, provide a way to integrate intermediary characteristics from different network stages. The MPRV-Net stacking approach accesses each level of the input image. The final output can be split into three separate, non-overlapping phases: the initial stage uses four patches, the subsequent stage uses two patches, and the third phase uses the initial picture.

Every stage C includes a reduced input image that is added to the residual image R_c that the suggested model predicts, which producesX_C:

$$X_C = I + R_C \tag{1}$$

The MPRV-Net's optimization process uses the following loss equation from beginning to end:

$$L = \sum_{C=1}^{3}\left[L_{char}(X_C, Z) + \lambda L_{edge}(X_C, Z)\right] \tag{2}$$

Where Z denotes the ground truth picture and $L_{char}$is the char bonnier loss.

$$L_{char} = \sqrt{\|X_C - Z\|^2 + \varepsilon^2} \tag{3}$$

Using an empirically adjusted constant ε of 〖10〗^(-3) for every experiment. Furthermore, L_{edge}often known as the edge loss, is defined as:

$$L_{edge} = \sqrt{\left\|\Delta(X_C) - \Delta(Z)\right\|^2 + \varepsilon^2} \tag{4}$$

Where the Laplacian operator is denoted by Δ. The value in equation (2), which is fixed at 0.05, controls the relative significance of the two loss variables.

Feature Extraction

Feature extraction is the procedure of transforming raw data into numerical characteristics that can be processed while preserving the information included in the original data set. MT-CNN algorithm is used as a feature extraction technique in the proposed system.

MT-CNN algorithm

Deep learning-based face recognition and alignment method MTCNN uses a cascading sequence of CNN to locate and identify faces in digital pictures or videos. There are three phases in the MT-CNN [20] model such as proposal network, refined network and output network.

i. Phase 1: Proposal network

The image data initially undergoes three convolution layers and one max pool layer operation once it enters the proposal network, transforming the original 12 ∗ 12 ∗ 3 image data into a 1 ∗ 1 ∗ 32 feature image. Using a set of 32 feature images 1 ∗ 1, two feature images can be generated for face classification, four feature images for regression frame judgment, and ten feature images for facial landmark localization. Face judgment and preliminary recommendations for face areas are carried out using face classification, bounding box regression, and face landmark positioning. The Refined Network receives this data as input for additional processing.

ii. Phase 2: Refined network

Prior to sending the picture data to the Refine Network, set the input image size to 24 ∗ 24 ∗ 3. This network creates 3 ∗ 3 ∗ 64 feature images using 3 convolutions and 2 softmax pools layer. It then creates 128 vectors using a fully connected layer. Two vectors can be produced for face categorization out of the 128 vectors in the complete connection layer, for regression box judgment, four vectors can be generated; for face contour point judgment, ten vectors can be generated. The prediction results are then further refined for NMS (non-maximum suppression) and the chosen bounding box regression. The output network receives this data for additional processing.

iii. Phase 3: Output network

Prior to sending the picture data to the output network, set the input image size to 48 * 48 * 3. This network generates 3 * 3 * 128 feature vectors using 4 convolution layers and 3 softmax pooling layers. It then uses the fully connected layer to generate 256 vectors. Two vectors are generated for face classification using the 256 vectors of the fully connected layer, four vectors are utilized for regression frame judgment, and ten vectors are used for the judgment of face contour points. Additionally, the output network is a sizable fully-connected layer made up of 256 vectors that preserves more picture information. It then uses this layer to carry out facial landmark localization, bounding box regression, and face identification before producing the coordinates of the face region's upper left corner. The five defining features are synchronized. Figure 2 shows the Architecture of MT-CNN algorithm.

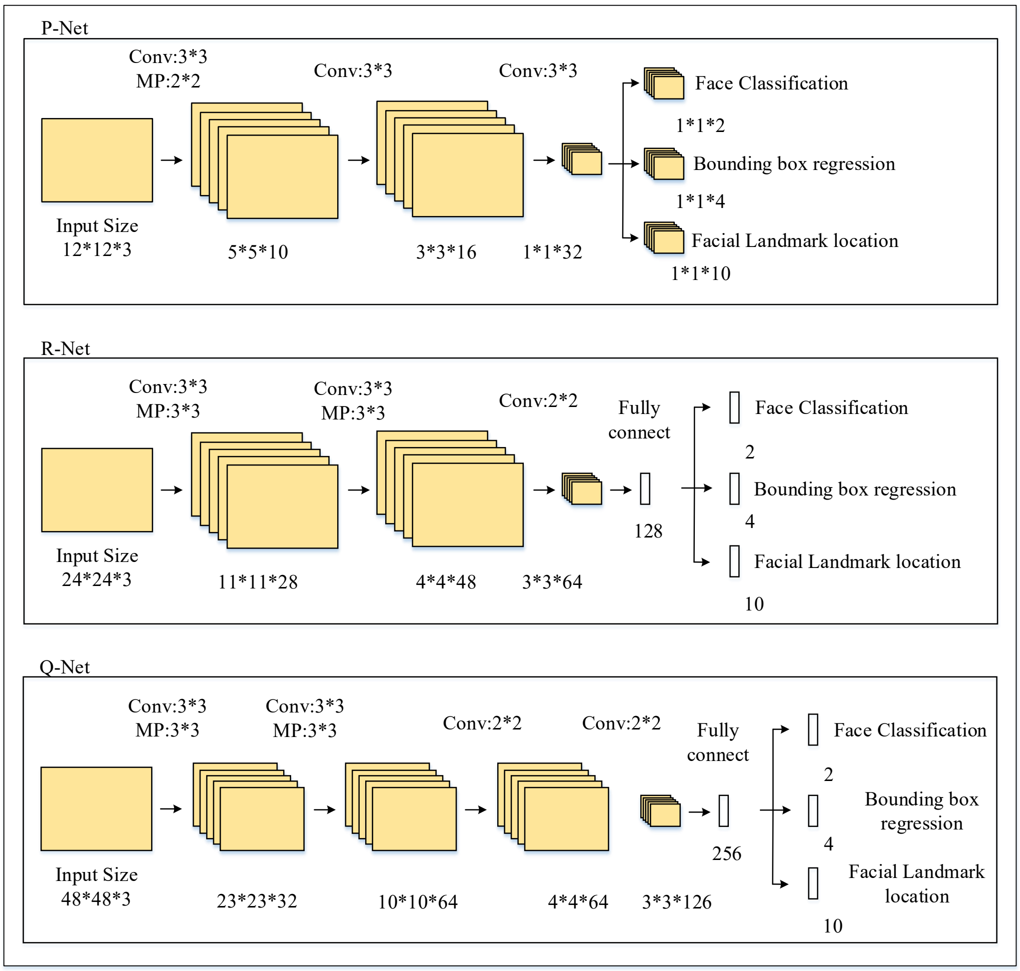

Figure 2. Architecture of MT-CNN algorithm.

Bounding box regression, landmark position, and face or non-face classifiers are the three subtasks of the MTCNN model. Firstly, it is possible to assess the distribution similarity between the training value and the true distribution since the face classification is part of a cross entropy cost function. The issue of the mean square loss function's learning rate dropping can be avoided by using the gradient descent method.

$$T_x^{det} = \left(b_x^{det}\log\left(q_x\right) + \left(1 - b_x^{det}\right)\left(1 - \log\left(q_x\right)\right)\right) \tag{5}$$

Where, q_x is found in the CNN network and denotes the likelihood that the sample is a face. b_x Indicates that the tag's face data has a value of either 0 or 1, where 0 denotes a non-human face and 1 denotes a face. Regression loss function for bounding box The Euclidean distance is used to calculate the value of a linear function known as regression. It is quite easy to calculate, and even after various representation domain modifications, the feature qualities remain the same.

$$T_x^{box} = \left\|\widehat{b_x}^{box} - b_x{}^{box}\right\|_2^2 \tag{6}$$

The formula above demonstrates that among them, $\widehat{b_x}^{box}$ box is a thinking vector and receives the return target from CNN. By using the same loss function as bounding box regression, facial landmark localization increases the model's accuracy by determining the minimal distance.

$$T_x^{iandmakr} = \left\|\widehat{b_x}^{iandmakr} - b_x{}^{iandmakr}\right\|_2^2 \tag{7}$$

Suppression of elements that are not maximum, or local maximum search, is known as NMS.

The neighbourhood's size and dimensions are the two variable characteristics in a local representation of the area. In face detection, the area chosen by the search algorithm is extracted by the feature, and the feature input classifier then determines the classification confidence. To choose the region projected to have the highest face score in the neighbourhood for additional detection, a non-maximum suppression method is required due to the overlap between the selected area and other selected regions. This reduces computing complexity.

Classification

In the supervised machine learning process of classification, the model attempts to predict the correct label of a given input set of data. Prior to being utilised to make predictions on fresh, unobserved data, the model in classification is thoroughly trained using the training set and then assessed using test data.

NASnet algorithm

Idea of an ideal network has been introduced by the deep neural network, taking it a step further [21]. Used a reinforcement learning-based strategy. The neural network architecture is modified by the parent AI after it evaluates the kid AI's effectiveness. To increase the network's efficiency, a lot of adjustments were made to the number of layers, weights, regularization techniques, etc. Using a reinforced evolutionary algorithm, the best candidates can be chosen. An algorithm for tournament selection is used to remove the cell with the lowest performance. Reinforcement mutations are carried out and the kid fitness function is optimized.

A set of blocks is called a cell in the NASnet architecture, and a block is the smallest unit. NASnet factorized the network into cells, which were then separated into blocks, to provide a search space. These blocks and cells are optimized for the chosen dataset; their quantity and type are not fixed. A block can perform a variety of operations, such as identity mapping, max-pooling, average-pooling, separable-convolutions, and normal convolutions. Two inputs are mapped by the blocks to a single feature map output. It requires adding of the elements one by one. Figure 3 shows the Architecture of NASnet algorithm.

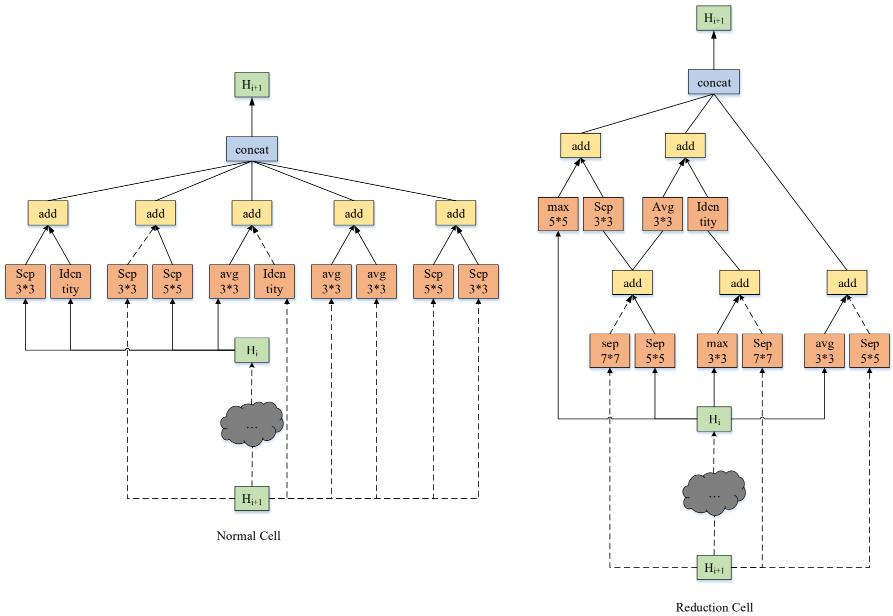

Figure 3. Architecture of NASnet algorithm.

When a feature map is contained in a cell, the block that cell receives will have the same dimensions as the characteristic map. If there are two strides, the size will be cut in half. A cell combination that works best is created. Three factors are the focus of network development: the cell shape, the number of filters in the first layer (F), and the number of cells to be stacked (N). After that, by modifying N and F in the first layer, the depth and width of the network can be managed. Several scaled models are built to fit the datasets after the search is complete. After that, the cells are joined in an ideal way to form the NASNet architecture. Pairwise combinations are used to create the hidden layers, which are then updated using concatenation. A series of pooling and convolution techniques can be applied to hidden layers.

In NASNet architecture, only the finest cells are chosen, as opposed to scanning through all of the cells. As a result, the search will be quicker and more generalist features may be found. The input feature map's height and breadth are preserved by the normal cell and are reduced by two by the reduction cell. Normal cells are layered between reduction cells to form the NASnet design. To produce extraordinary accuracy NASnet Large has taken N to be 6, whereas NASnet Mobile was primarily concerned with functioning on a constrained budget. The chosen set of steps reduces the input size of 224 × 224 × 3 to an output size of 7 × 7 by using 5B cells for both the reduction and normal cells. In NASnet, a novel idea known as scheduled drop path is presented. As the network is trained, each path in the cell is dropped with a probability that increases linearly. The reduction cell states that various procedures are carried out on the Hi−1 and Hi−2 cells prior to their concatenation in order to construct the Hi. A series of steps on Hi−1 and Hi−2 cells, and finally their union. These prediction steps are repeated B times by the RNN. The NASNet architecture significantly increased operation speed. As one moved to more recent architectures, a situation where fewer learnable parameters were present was noticed. It has been noted that as CNN models have evolved, each architecture has added special features. Convolution and max-pooling were the sole processes utilized in the beginning. With the introduction of the Inception module, spatial dimension was given more consideration. CNN models have to contend with the threat of vanishing gradient during that period. ResNet made a commendable suggestion for leveraging identity mapping to solve this issue. Mobile Net developed the idea of separable-convolution to lower the number of learnable parameters. Pseudocode for the proposed approach is given in algorithm 1.

Algorithm 1: Pseudocode for criminal facial recognition

Input*: image related to criminal and non-criminal faces*
begin
 # performing pre-processing on the input image (I)
{
P=prepro(I)
{
PR=Multi stage progressive V-Net
}
 # perform feature extraction (PR)
{
FE= MTCNN (PR)
}
#classifying features (FE)
{
Cl= NasNet(FE)
}
}
Output: criminal faces are accurately identified

3. RESULT AND DISCUSSION

NASnet algorithm is developed for classifying criminal and non-criminal. Initially the data is collected as images. Then the images are pre-processed by Multi stage progressive V-net algorithm. After the image is feature extracted by MT-CNN algorithm. Finally, the image is classified by NASnet algorithm. The testing is carried out using Python 3.8 combining CPU, a 16-bit operating system, an Intel Core i5, an NVidia GeForce GTX 1650, and 16GB of RAM. The Parameters for NASnet algorithm is shown in Table 1.

Table 1. Parameter of NASnet algorithm

Parameters	Range
Learn rate	0.00001
momentum	0.8
Batch size	32
Train layers	All
Optimizer	Adam
Activation function	ReLU

Table 1 contains the parameters of NASnet classifier which are learning rate, momentum, batch size, train layers, optimizer and activation function. Table 2 shows the Sample pre-processed and segmented images.

Table 2. Sample pre-processed and segmented images

Original image	Pre-processed image	Segmented image	Output image
			
			
			
			
			

Dataset Description

The data was collected from open source Kaggle [22] [23]. The dataset which contains the facial image. The CK+ and fbi.gov websites are used in the designed model to obtain the dataset. A total of 934 picture data points are collected and categorized as either criminal or non-criminal. 339 of the photos in this collection are considered criminal, while 595 are not. Form the entire data, 80% of the data taken for training and 20% of the data was taken for testing.

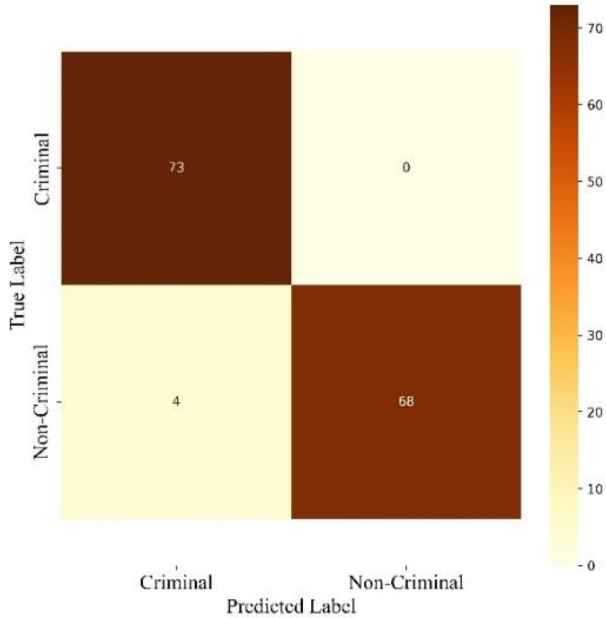

Figure 4. Confusion matrix.

Confusion matrix analysis is one method for evaluating the accuracy of a classification system. Proposed approach's confusion matrix plot is shown in Figure 4. A confusion matrix is a table that usually shows how well a classification model performs on a set of test data with known real values. Basic predictive metrics such as recall, specificity, precision, and accuracy are displayed via confusion matrices. This demonstrates how the confusion matrix's intended and real labels are interpreted incorrectly. In that sequence, the anticipated values for criminal and non- criminal 73 and 68.

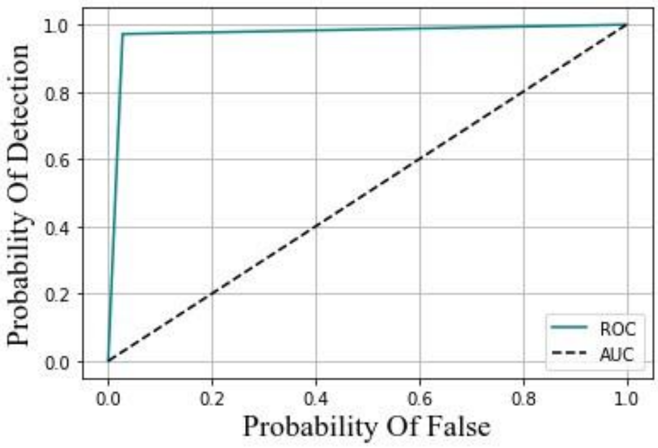

Figure 5. ROG matric's.

The ROC Curve is shown in Figure 5. Receiver operating characteristic curves, or ROC curves, are graphs that display a classification model's performance over all categorization thresholds. The True Positive Rate and False Positive Rate are the two parameters plotted on this curve. The system becomes more separable as a result of the curve that approaches 1.

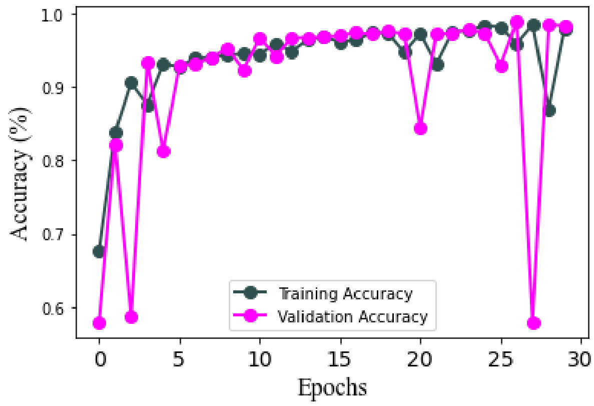

Figure 6. Evaluation of accuracy and Epochs in Training and validation.

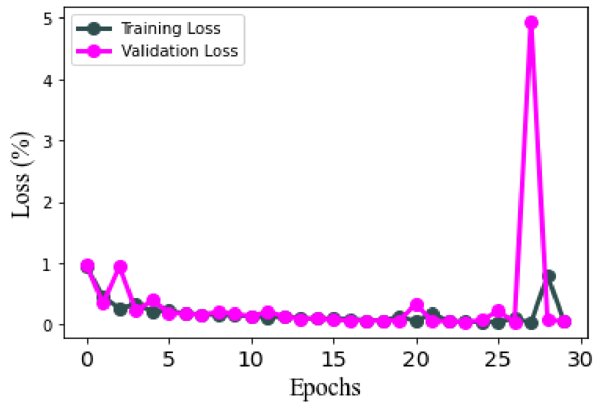

Figure 7. Evaluation of loss and Epochs in Training and validation.

The comparison of accuracy by varying the Epochs is shown in the figure 6. Accuracy is the degree to which the values in the observational table or the outputs on the graph are closely spaced from one another, allowing for a comparison of the results. The percentage of correctly classified positive answers divided by the total number of actually positive answers is known as accuracy. Figure 7 compares the loss by varying the Epochs. Plotting the trade-off between a classifier's loss and Epochs is known as the Receiver Operating Characteristic (ROC) curve.

Comparison Analysis

Graphs given below compares the acc. Each graph is compared with various current algorithms such as Accuracy, F1_Score, false discovery rate (FDR), Miss Rate (MR), Fall-out**,** False omission rate **(**FOR), Negative positive rate (NPV), Selectivity, Hit Rate, Error and PPV. Each graph is compared with various current algorithms such as diffuse convolutional neural network (DCNN), ResNet-101, convolutional long short-term memory (CLSTM), internal externals- convolutional neural network (IE-CNN) and the proposed system is NASnet.

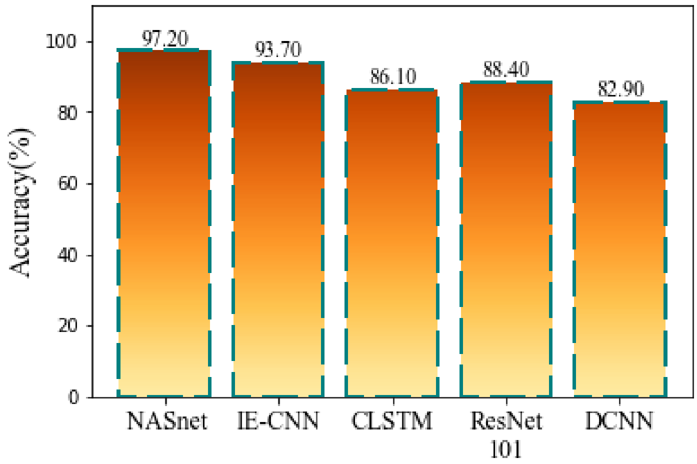

Figure 8. Comparison of Accuracy.

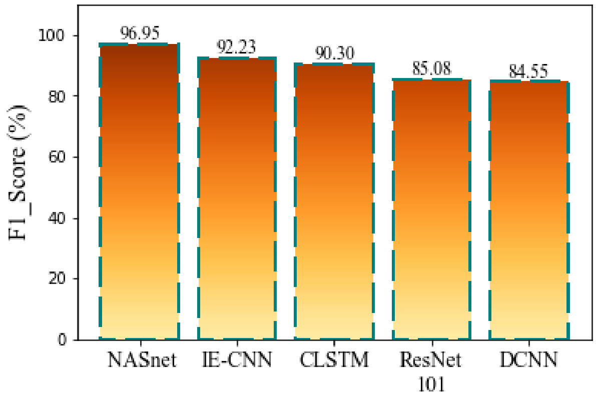

Figure 9. Comparison of F1_Score.

The Accuracy parameter for the suggested and current methods is contrasted in Figure 8. The Accuracy value of NASnet is 97.20% for the suggested approach was found to be greater than the values of 93.70%, 86.10%, 88.40% and 82.90% for the IE-CNN, CLSTM, ResNet and DNN approaches, respectively. Figure 9 shows an analysis using F1_Score metrics comparing recommended and existing approaches. The F1_Score percentages obtained by the current methods IE-CNN, CLSTM, ResNet and DNN are 92.23%, 90.30%, 85.08% and 84.55%, in sequence. Yet, the F1_Score value of the suggested NASnet model is 96.95%.

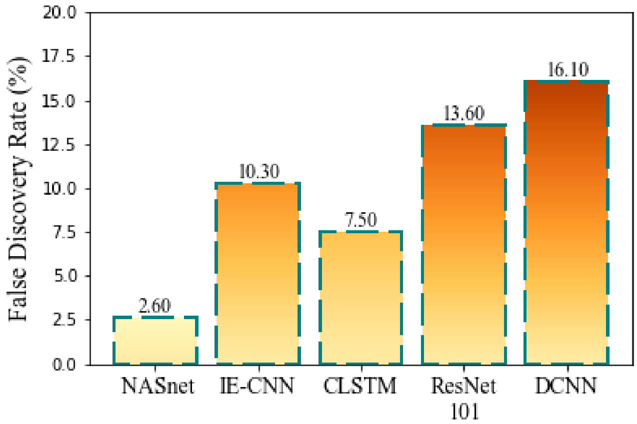

Figure 10. Comparison of FDR.

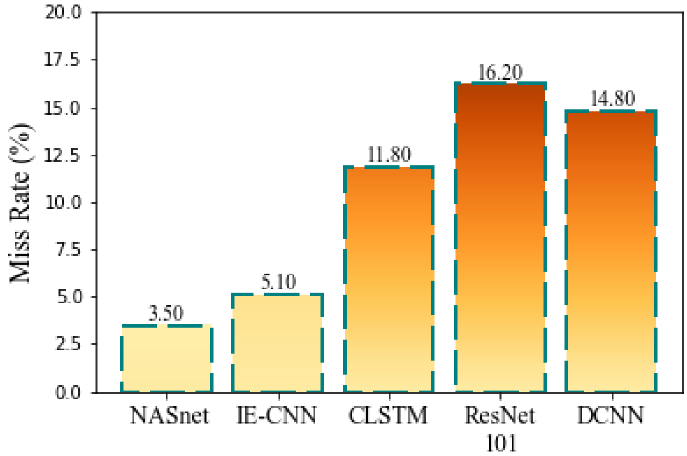

Figure 11. Comparison of MR.

The graph shown in figure 10 presents a FDR analysis of four established methods IE-CNN, CLSTM, ResNet and DNN along with a proposed approach NASnet. The FDR of proposed NASnet is 2.60%, and the existing approaches as LSTM are 10.30%, DBN is 7.50%, RNN is 13.60% and FNN is 16.10%. The FDR percentage of NASnet increased than the existing system. Figure 11 displays the MR for the proposed and existing approaches. With an ensemble learning classifier, the suggested NASnet approach has a MR of 3.50%. On the other hand, the MR values for the IE-CNN, CLSTM, ResNet and DNN algorithms that are currently in use are, respectively, 5.10%, 11.80%, 16.20% and 14.80%. It suggests that the suggested method outperforms the current ones.

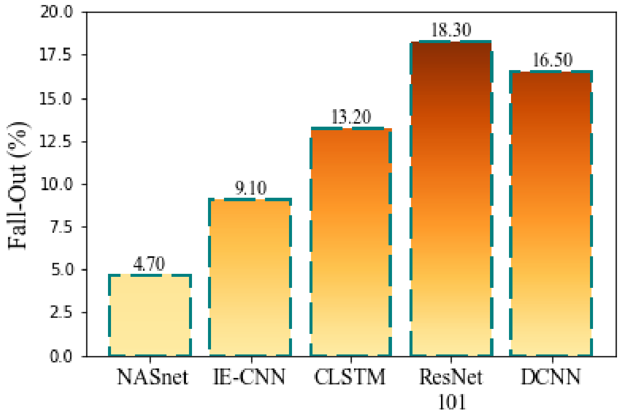

Figure 12. Comparison of Fall-out.

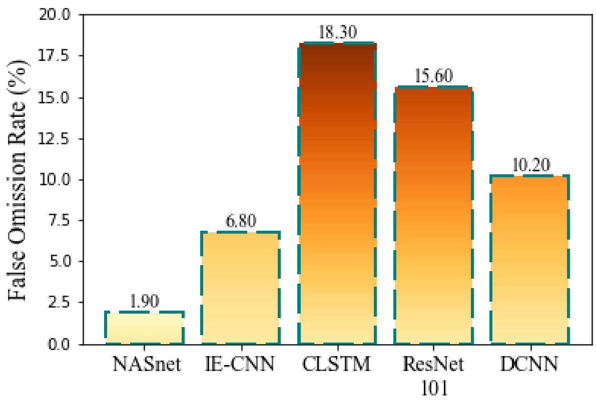

Figure 13. Comparison of FOR.

Figure 12 shows the Fall-out comparison of NASnet, LSTM, DBN, RNN and FNN. The Fall-out of the proposed system NASnet increases than all the existing systems. The Fall-out percentage of NASnet is 4.70%, LSTM is 9.10%, DBN is 13.20%, RNN is 15.60% and FNN is 10.20%. The Fall-out of existing systems IE-CNN, CLSTM, ResNet and DNN decreased than the proposed system NASnet. The FOR metrics of the suggested and current methods are contrasted in Figure 13. The FOR percentage of the other systems IE-CNN, CLSTM, ResNet and DNN is 1.90%, 6.80%, 18.30%, 15.60% and 10.20% for the proposed NASnet system. The suggested system NASnet FOR percentage is increases than the Existing system.

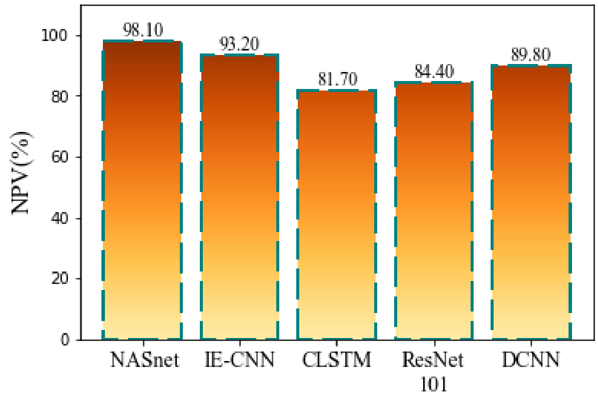

Figure 14. Comparison of NPV.

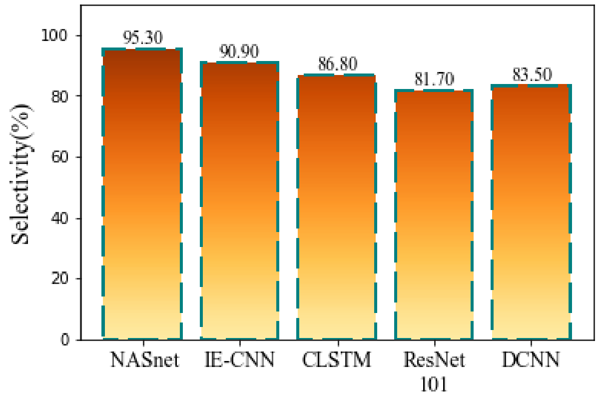

Figure 15. Comparison of Selectivity.

The graph shown in figure 14 presents an NPV analysis of four established methods IE-CNN, CLSTM, ResNet and DNN along with a proposed approach NASnet. The NPV of proposed NASnet is 98.10%, and the existing approaches as LSTM is 93.20%, DBN is 81.70%, RNN is 84.40% and FNN is 89.80%. The NPV percentage of NASnet increased than the existing system. Figure 15 shows the comparison of Selectivity by the proposed NASnet and the existing IE-CNN, CLSTM, ResNet and DNN. The x-axis represents different existing and proposed approaches, while the y-axis indicates the Accuracy percentage of the existing and proposed approaches. The Selectivity of NASnet is 95.30% and higher in the existing system. The Selectivity of existing approaches are 90.90% for LSTM, 86.80% for DBN, 81.70% for RNN and 83.50% for FNN.

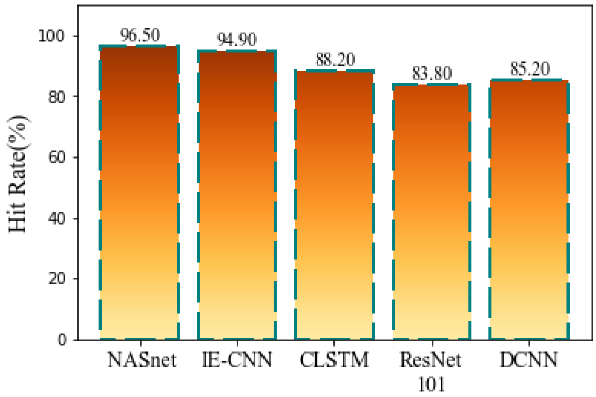

Figure 16. Comparison of Hit Rate.

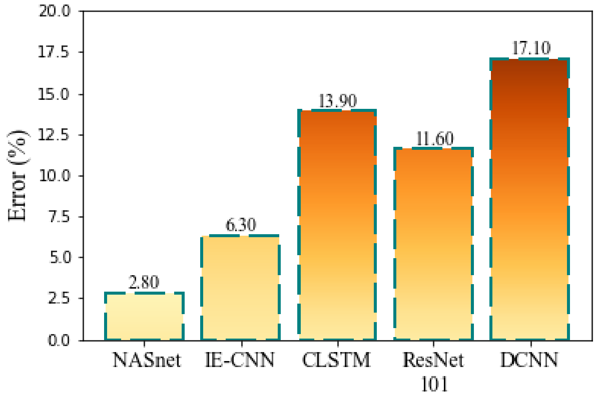

Figure 17. Comparison of Error.

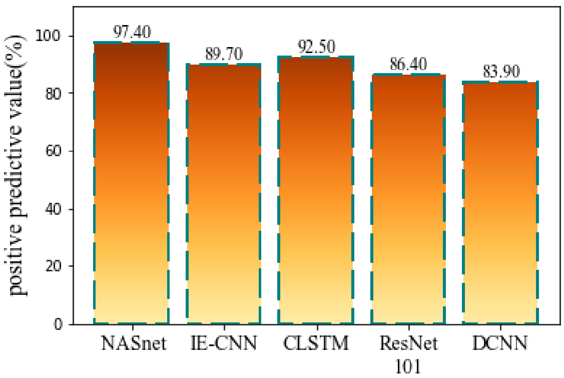

Figure 18. Comparison of PPV.

Figure 16 shows the comparison of the Hit Rate by the proposed NASnet and the existing IE-CNN, CLSTM, ResNet and DNN. The x-axis represents different existing and proposed approaches, while the y-axis indicates the Hit Rate percentage of the existing and proposed approaches. The Hit Rate of NASnet is 96.50% and increases in the existing system. The Hit Rate of existing approaches are 94.90% for LSTM, 88.20% for DBN, 83.80% for RNN and 85.20% for FNN. The graphical representation of the Error comparison parameter is shown in Figure 17. The error value of the proposed NASnet is 2.80% and existing approaches are LSTM is 6.30%, the DBN is 13.90%, the RNN is 11.60% and the FNN is 17.10%. The error percentage of the proposed decreases than the existing system. The graphical representation of the PPV is shown in Figure 18. The value of 97.40% is found in the comparison of the present and previous techniques, such as the NASnet. The values of the LSTM are 89.70%, the DBN is 92.60%, the RNN is 86.40% and the FNN is 83.90%.

4. CONCLUSION

Deep learning based NASnet algorithm is developed for recognizing the criminal faces. Face recognition can extract the individuality features of the human face. Human face detection and recognition play a significant part in applications such as face picture database management and video surveillance. One straightforward and adaptable biometric technique is face recognition. Face recognition technology is used in security, healthcare, criminal justice, and other fields where the ability to recognize people is essential. Technology advancements have made it easier to extract facial traits from people. Number of offenders and the crime rate have both increased unusually, which creates grave concerns about security-related issues. Police officers primarily deal with crime prevention and criminal identification since safeguarding people's property and lives is their first concern. However, there are not always enough police officers to combat crime. So, automated identification is developed to recognizing the criminals. The images are pre-processed using the multi-stage progressive V-net approach. This approach, which is divided into three steps, uses V-net architectures to learn every feature in the image. Next, the MT-CNN algorithm, which produces 62 facial landmarks, is used to align the improved image for facial features. Finally, by training and evaluating the retrieved facial features using the MT-CNN, the NASnet architecture is employed to predict the criminal face. Furthermore, the suggested approach's efficacy was confirmed and assessed in comparison to other cutting-edge techniques such as IECNN, CLSTM, ResNet 101, and DCNN. The proposed approach has 97.2% of accuracy, 96.9% of f1-score, 98.1% of NPV, 97.4% PPV, 95.30% of selectivity, 1.90% of FOR, 4.70% of fall-out, 2.80% of error, 3.50% miss rate and 2.6% FDR. This graphical illustration demonstrates that the outcomes of the suggested approach are substantially better compared to those of the existing methods. The results showed that, when compared to alternative approaches, the suggested strategy can yield better outcomes by resolving the issue of lower accuracy and time consumption.

5. LIMITATION AND FUTURE WORK

The suggested method is applied to criminal face detection and can readily handle more complicated tasks, handle different kinds of data, and yield more precise and effective outcomes. However, these models can result in overfitting, where the model works well on training data but not for unknown data. And it has large network size and complexity. So, future work focussed on incorporating an optimization algorithm with this proposed model to avoid overfitting problems. In the future, incorporate the alarms into the system for criminal detection. The system will only alert those who are not present to monitor the CCTV room when a match is made, letting them know that someone has been located in that public area using the database

ACKNOWLEDGEMENT

Funding. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Conflict of Interest. The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Availability of data and material. Not applicable

Code availability. Not applicable

Author contributions. The corresponding author claims the major contribution of the paper including formulation, analysis and editing. The co-author provides guidance to verify the analysis result and manuscript editing.

Compliance with ethical standards. This article is a completely original work of its authors; it has not been published before and will not be sent to other publications until the journal’s editorial board decides not to accept it for publication.

REFERENCES

[1] Kortli, Y., Jridi, M., Al Falou, A., & Atri, M. (2020). Face recognition systems: A survey. Sensors, 20(2), 342.

[2] Petrescu, R. V. (2019). Face recognition as a biometric application. Journal of Mechatronics and Robotics, 3, 237-257.

[3] Kukharev, G. A., & Kaziyeva, N. (2020). Digital facial anthropometry: application and implementation. Pattern Recognition and Image Analysis, 30, 496-511.

[4] Andrejevic, M., & Selwyn, N. (2020). Facial recognition technology in schools: Critical questions and concerns. Learning, Media and Technology, 45(2), 115-128.

[5] de Bruijn, H., Warnier, M., & Janssen, M. (2022). The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. Government information quarterly, 39(2), 101666.

[6] Baughman, S. B. (2021). Crime and the Mythology of Police. Wash. UL Rev., 99, 65.

[7] Taskiran, M., Kahraman, N., & Erdem, C. E. (2020). Face recognition: Past, present and future (a review). Digital Signal Processing, 106, 102809.

[8] Zhang, J. (2021). Music feature extraction and classification algorithm based on deep learning. Scientific Programming, 2021(1), 1651560.

[9] Schiks, J. A., van de Weijer, S. G., & Leukfeldt, E. R. (2022). High tech crime, high intellectual crime? Comparing the intellectual capabilities of cybercriminals, traditional criminals and non-criminals. Computers in Human Behavior, 126, 106985.

[10] Zhao, Y., Li, T., Zhang, X., & Zhang, C. (2019). Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future. Renewable and Sustainable Energy Reviews, 109, 85-101.

[11] Hashemi, M., & Hall, M. (2020). RETRACTED ARTICLE: Criminal tendency detection from facial images and the gender bias effect. Journal of Big Data, 7(1), 2.

[12] Sandhya, S., Balasundaram, A., & Shaik, A. (2023). Deep Learning Based Face Detection and Identification of Criminal Suspects. Computers, Materials & Continua, 74(2).

[13] Sakiba, C., Tarannum, S. M., Nur, F., Arpan, F. F., & Anzum, A. A. (2023). Real-time crime detection using convolutional LSTM and YOLOv7 (Doctoral dissertation, Brac University).

[14] Ahire, N., Dachawar, A., Disale, P., Galande, S., & Gawande, N. (2023, August). A Comparative Study on Facial Detection for Criminal Probe. In 2023 3rd Asian Conference on Innovation in Technology (ASIANCON) (pp. 1-7). IEEE.

[15] Jahan, N., Bhuiyan, P. K., Moon, P. A., & Akbar, M. A. (2020, July). Real time face recognition system with deep residual network and KNN. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 1122-1126). IEEE.

[16] You, S. (2021). Discrete Lion Swarm Optimization Algorithm for face recognition: Discrete Lion Swarm Optimization Algorithm for face recognition. Multimedia Research, 4(3).

[17] Patil, G., Bethu, S., Bhavani, Y. D., Harikrishna, B., Vijayalakshmi, A., & Babu, B. S. (2022, August). IoT and Cloud based Face detection application design for Surveillance systems using Deep Learning. In Proceedings of the 2022 6th International Conference on Cloud and Big Data Computing (pp. 49-55).

[18] Kulkarni, N., Mantri, D., Pawar, P., Deshmukh, M., & Prasad, N. (2022, October). Occlusion and spoof attack detection using Haar Cascade classifier and local binary pattern for human face detection for ATM. In AIP Conference Proceedings (Vol. 2494, No. 1). AIP Publishing.

[19] Haque, I. R. I., & Neubert, J. (2020). Deep learning approaches to biomedical image segmentation. Informatics in Medicine Unlocked, 18, 100297.

[20] Zhang, N., Luo, J., & Gao, W. (2020, September). Research on face detection technology based on MTCNN. In 2020 international conference on computer network, electronic and automation (ICCNEA) (pp. 154-158). IEEE.

[21] Qin, X., & Wang, Z. (2019). Nasnet: A neuron attention stage-by-stage net for single image deraining. arXiv preprint arXiv:1912.03151.

[22] Dataset 1: https://www.fbi.gov/wanted/fugitives

[23] Dataset 2: https://paperswithcode.com/dataset/ck