Journal of Artificial Intelligence and Capsule Networks (ISSN: 2582-2012)
www.irojournals.com/aicn/

Deciphering Crime Suspects' Emotion for Investigative Precision using Deep Learning

Harikrishnan A[1], Nikhil Suresh[2], Sanil M[3], Sumesh P[4], Remya M[5]

Computer Science, Jawaharlal College of Engineering and Technology, KTU, Palakkad, India

Email: [1]harikrish2002ba@gmail.com, [2]nikhilsuresh765@gmail.com, [3]sanilm4637@gmail.com, [4]sumeshpkv3@gmail.com, [5]remya3956.cse@jawaharlalcolleges.com

Abstract

Using Convolutional Neural Networks (CNNs) for Facial Emotion Recognition (FER) in criminal investigations is the main goal of the proposed study. Using cutting-edge CNN architectures trained on a variety of datasets to precisely identify emotions including joy, anger, fear, surprise, disgust, and sadness is one of the main components. The research seeks to improve the effectiveness of investigations by giving law enforcement a tool to interpret the subtle emotional cues in the facial expressions of potential suspects in crimes. The technique has potential uses in human-computer interface, security, and surveillance in addition to criminal investigations. The larger significance is in helping to create more precise and effective investigative procedures, which will ultimately promote a society that is safer and more secure. The study focuses on providing a brief overview, with a suggested method utilizing CNNs to trace crime suspects.

Keywords: CNN, Facial Emotion Recognition, Deep Learning, Artificial Intelligence, Facial Recognition, Image Processing

1. Introduction

The incorporation of cutting-edge technologies has become essential in today's criminal investigation landscape to improve accuracy and efficiency. In the context of criminal investigations, this research offers a novel method for Facial Emotion Recognition (FER) using Convolutional Neural Networks (CNNs) [9, 11]. The research aims to utilize the most advanced

DOI: https://doi.org/10.36548/jaicn.2024.2.002
Received: 08.03.2024, received in revised form: 03.04.2024, accepted: 25.04.2024, published: 04.05.2024

CNN architectures that have been trained on a variety of facial expression datasets, with a specific focus on the nuanced interpretation of the emotional states of crime suspects [12]. The range of emotions that are addressed includes happiness and anger as well as fear, surprise, disgust, and sadness. The potential of this research is to transform investigative procedures and give law enforcement organizations a cutting-edge tool to understand the psychological characteristics of criminal suspects. During interrogations, detectives can make well-informed decisions by interpreting the facial expressions of suspects [10]. The suggested system has potential applications in areas other than its principal use, such as security, surveillance, and human-computer interaction [14-15]. This novel method is characterized by ethical concerns and a dedication to eliminating potential biases. These elements provide responsible and fair deployment in the aim of a better and more secure society. The incorporation of state-of-the-art instruments into law enforcement methods is a crucial step in promoting efficiency and accuracy in criminal investigations, particularly as technology advances [16].

2. Related Work

Modi, Shrey, and Mohammed Husain Bohara.(2021) [1].

The research proposes a strategy employing convolutional neural networks (CNNs) trained on the FER2013 dataset to identify facial expressions, achieving an accuracy of 92.8% on the test set. The process consists of four phases: data preprocessing, feature extraction, categorization, and evaluation. Data preparation involves scaling the photos to 48x48 pixels, standardizing them, and converting them to grayscale. Feature extraction is performed using a CNN architecture with two convolutional layers, two pooling layers, and two fully connected layers. The images are then classified into six emotions using a softmax classifier. The CNN model outperforms alternative methods, with an accuracy of 92.8%, compared to 88.2% for the SVM and 86.6% for the random forest classifier.

Jaiswal, Akriti, A. Krishnama Raju, and Suman Deb. (2020) [6].

The Japanese Female Facial Expression (JAFFE) and Facial Expression Recognition Challenge (FERC-2013) datasets are used in the study to demonstrate a deep learning-based facial emotion identification technique. For CNN feature extraction, the model is composed of two submodels that are flattened into vectors, concatenated into a vector matrix, and sent to a fully connected layer for analysis. The last output layer makes classification possible. The model consists of flatten layers, maxpooling, local contrast normalization, and a convolutional layer with 64 filters. The model can then be coupled to a softmax output layer that can identify seven emotions by concatenating it with two other models that are similar to it. Using the FER dataset for facial emotion detection and the JAFFE dataset for face emotion detection, the model's average validation accuracy was 70.14 percent and 98.65 percent, respectively.

Singh, Shekhar, and Fatma Nasoz.(2020)[2].

It suggests a unique convolutional neural network (CNN) architecture for face expression recognition. Using the FER2013 dataset, the suggested model obtains a test accuracy of 61.7% without the need for any feature extraction or pre-processing methods. Six convolutional layers, two fully connected layers, three max-pooling layers, and an output layer with a softmax activation function make up the model. The 35,887 face crops in the FER2013 dataset—which includes training, validation, and testing images with 28,709, 3,589, and 3,589, respectively—are used by the authors. Every image is in grayscale and has a resolution of 48 x 48 pixels. The authors accomplish light correction via histogram equalization and employ a Haar Cascade classifier for face detection. In order to avoid over fitting, the authors additionally apply a dropout of 0.2 to the fully linked layers.

Begaj, Sabrina, Ali Osman Topal, and Maaruf Ali. (2020)[8].

The study on convolutional neural networks' (CNNs) use for facial emotion recognition is presented in this publication. In order to ascertain the best efficient method for identifying the seven fundamental emotions—angry, contempt, disgust, fear, pleasure, sadness, and surprise—the authors experiment with several CNN architectures and datasets. With an accuracy of 86.64% on the iCV MEFED dataset, the authors discover that a CNN with four convolutional layers, four max pooling layers, one dropout layer, and two fully connected layers produces the best results. They also discover that the accuracy of the CNN may be increased by employing data augmentation methods such image flipping, scaling, and rotation. The authors come to the conclusion that CNN is a promising method for recognizing face emotions and that more data may be used to further enhance them.

Mehendale, Ninad.(2020)[4].

It suggests a cutting-edge method for convolutional neural networks (CNNs)-based facial emotion identification. Testing the FERC method on a dataset of 10,000 photos yields a 96% accuracy rate. The two primary phases of the FERC algorithm are the extraction of facial features and the elimination of backdrop. In the background removal stage, the face is separated from the background using a skin tone-based feature and the Hough transform for circles-in-circles filters. In order to extract facial features from the cropped face image, the facial feature extraction stage employs a two-part CNN. The CNN initially eliminates the image's backdrop before extracting facial features to create an expressional vector (EV). The EV is used to classify the image into one of five basic emotions: displeasure/anger, sad/unhappy, smiling/happy, feared, and surprised/astonished.

H. Zhang, A. Jolfaei and M. Alazab.(2019) [3].

Using image processing, edge computing and a convolutional neural network, a novel approach to face expression recognition is put forth. The process extracts the edge of each layer from the input facial expression image, normalizes it, and then superimposes the edge information on feature maps. It then reduces the dimension of the retrieved implicit features using a maximum pooling technique. A Softmax classifier is used for the classification and recognition of the test sample image. A simulation experiment was carried out to confirm the resilience of the strategy in the presence of complicated backdrops. With less repetitions and a training pace 1.5 times faster than the contrast algorithm, the findings demonstrated an average recognition rate of 88.56%.

Sang, Dinh Viet, and Nguyen Van.(2017) [5].

The proposed architectures, inspired by VGGNet, differ in depth from 8 to 14 layers, consisting of convolutional, pooling, and fully connected layers. The ReLU activation function is applied after each layer, except for the output layer. The output layer contains 7 neurons corresponding to emotional labels. Different loss functions are used in experiments, including cross-entropy and L2 multi-class SVM. BKVGG14 with L2 multi-class SVM achieves the best accuracy, achieving 71.2% on the FERC-2013 public test set and 72.2% on the private test set.

3. Proposed Work

Several essential elements are used to create a deep learning system that can accurately interpret the emotions of crime suspects for investigation purposes. First and foremost, gathering data is essential. This includes recording interviews on audio and video and adding physiological data from equipment like heart rate monitors. Pre-processing is required to clean and synchronize physiological, audio, and visual inputs during the data processing step. The next step is feature extraction, which transforms physiological reactions, voice modulation, and facial expressions into a format that can be used with machine learning models. Model Development is the system's key component. Training deep learning models on pre-processed data involves using convolutional neural networks (CNNs) for image processing. These models learn to identify and classify emotional states based on the multimodal data. Figure 1 illustrates the system architecture of the proposed system.

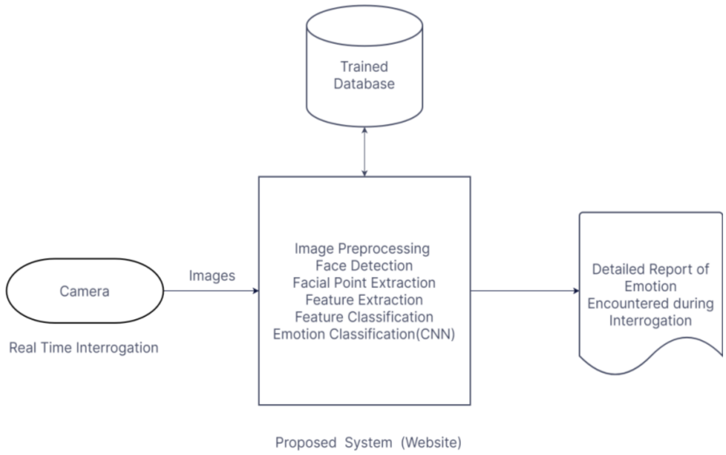

Figure 1. System Architecture

Image and Video Pre-Processing

Face detection algorithms are used to extract facial regions from images or video frames, ensuring proper alignment for consistent features. Pixel values are normalized to a standard scale, ensuring uniform intensity across images. Images are rescaled to a consistent size for efficient processing and model stability. If imbalanced datasets are present, data augmentation techniques are applied to artificially increase the dataset size.

Feature Extraction

Facial landmark detection involves identifying key facial landmarks like eyes, nose, and mouth for subtle expression changes. Deep learning-based feature extraction uses pre-trained convolutional neural networks (CNNs) to learn hierarchical features from facial images, capturing both low-level details and high-level abstract representations.

Emotion Classification

The Emotion Classification Model trains deep learning models to classify emotions based on extracted features, while Ensemble Learning (Optional) combines outputs from multiple models for improved accuracy.

User Interface

The Investigator Dashboard offers a user-friendly interface for investigators to view and interpret results, while alerts and notifications are provided to alert them of significant emotional cues or anomalies.

3.1 FER 2013 Dataset

A key resource in the field of facial emotion recognition (FER) research, the FER 2013 dataset offers an insight into the intriguing realm of interpreting emotions from facial expressions. Imagine thousands of faces that represent the range of human emotions, each uniquely deformed. This dataset is used to train algorithms to understand the complex language of facial expressions. A selection of sample photos from the used dataset are shown in Figure 2.

Figure 2. Dataset

It includes a training dataset, a public test dataset (which serves as our project's validation dataset), and an additional private test dataset (which is the same size as the public test dataset and will be used to assess prediction performance). 35,887 samples in an image set, with a training set of 80%, a validation set of 10%, and 10% of the test set. There are 28709 images for training, 3589 images for testing, and 3589 images for validation. Convert strings into integer lists. Resize to 48 by 48 and use 255.0 to normalize the grayscale image. Labels can be encoded one-hot, for example, class 3 to [0,0,0,1,0,0,0].

3.2 CNN

A specialized deep learning architecture called a convolutional neural network (CNN) is made for processing visual data, especially for image identification applications. Convolutional layers, which make up CNNs, use filters to extract local features such as edges and textures. Feature maps are produced by several filters, resulting in hierarchical representations. After that, pooling layers down sample these maps while keeping important details and cutting down on computational complexity. ReLU and other activation functions add non-linearity to the network to improve its ability to recognize complicated patterns. CNNs are particularly good in picture classification, object detection, and facial recognition because they perform well in situations where spatial hierarchies and local patterns are important. Their design, which was influenced by how the human brain processes images, has found extensive use in a variety of fields, demonstrating their versatility and effectiveness.

3.3 System Working

Figure 3 illustrates the flow diagram representing the systematic flow in the system.

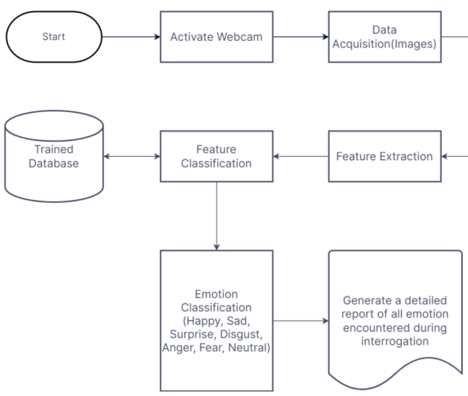

Figure 3. Flow Diagram

The process of emotion recognition involves a webcam being activated and capturing a live video stream of a person's face. The captured images are processed for the emotion recognition model, which includes tasks like landmark detection, facial feature extraction, and data normalization. Optional steps include saving points or detecting facial points to monitor subtle changes in expressions and identify areas associated with certain emotions. Feature extraction is done using geometric, textural, or motion-based features. Emotion classification is performed using a trained machine learning model, which outputs the classified emotion and a confidence score. For improved accuracy, some systems may employ multiple emotion classification models or approaches. A detailed report of all emotions encountered during interrogation is generated, summarizing the person's emotional state over time, including dominant emotions, emotional transitions, and statistical analyses.

3.4 Data Flow Diagram (DFD)

The data flow diagram level 0, also known as a context diagram, shows a system that receives real-time video input, consults a trained database, and outputs a report on various emotions. Figure 4 illustrates the level 0 DFD.

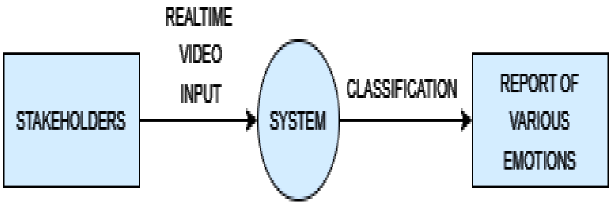

Figure 4. Level 0

Level 1 DFDs typically expand upon a level 0 diagram by showing the main sub-processes within the system. The system acquires real-time data, preprocesses it, extracts features, trains a model, evaluates the results, and selects the best model for final output. Figure 5 illustrates the level 1 DFD.

Figure 6 to 9 represents various levels of level 2 DFD:

Level 2.1 (Figure 6): The process selects data from camera feed, stores it in a raw data repository, and then pre-processes the data for further analysis.

Level 2.2 (Figure 7): The process acquires data from webcam, pre-processes it, and extracts features for face recognition.

Level 2.3 (Figure 8): The machine learning module pre-processes data, trains a model using a training dataset and algorithm, and evaluates the resulting classification models.

Level 2.4 (Figure 9): The process compares two trained models, analyzes the results, and selects the best model for classification.

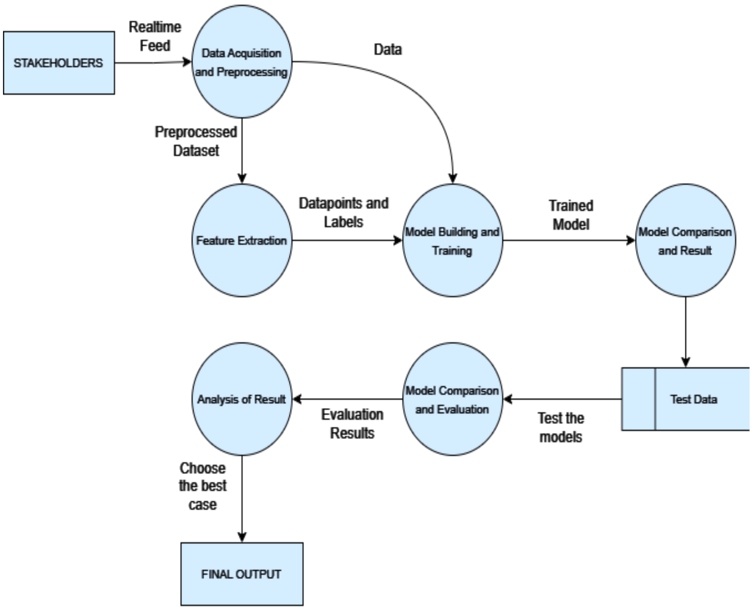

Figure 5. Level 1

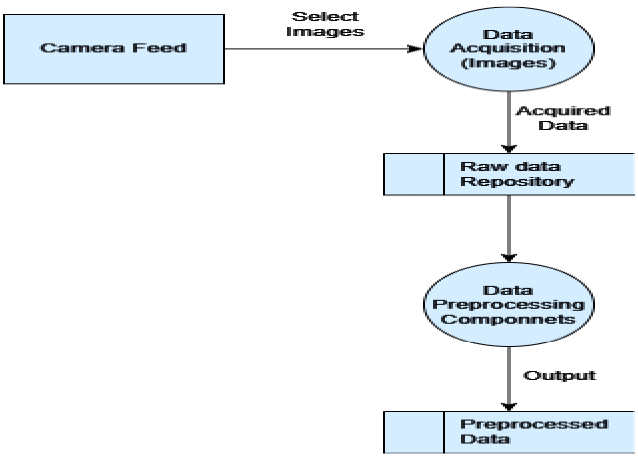

Figure 6. Level 2.1 (Data Acquisition and Pre-Processing)

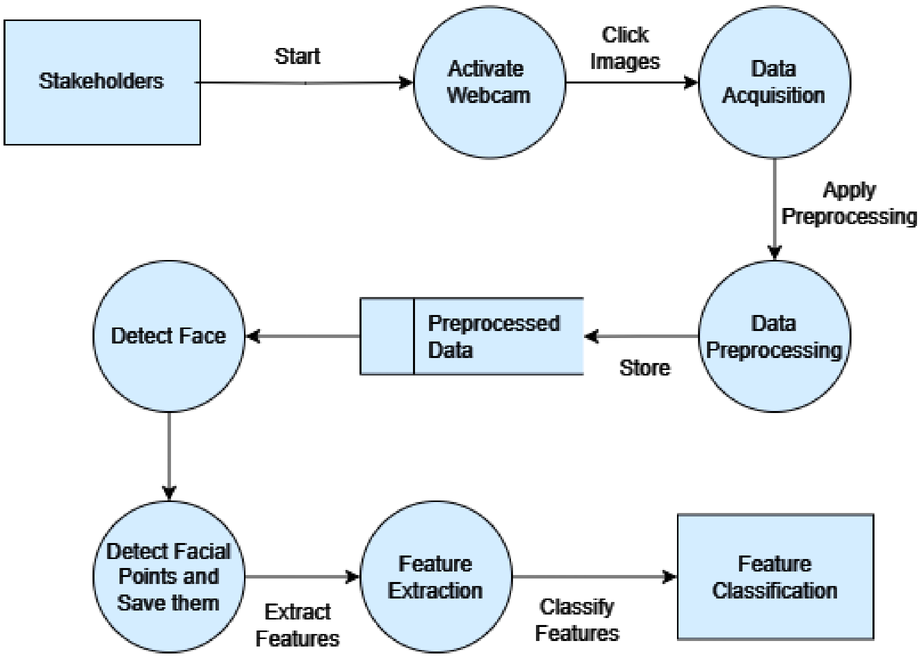

Figure 7. Level 2.2 (Data Acquisition and Pre-Processing)

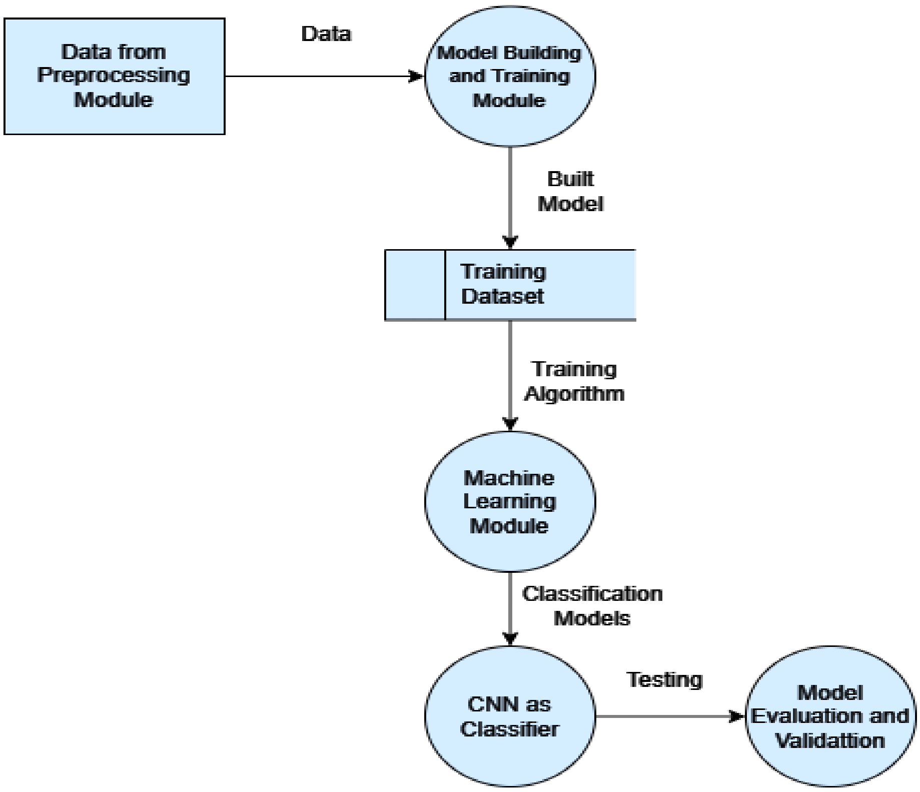

Figure 8. Level 2.3 (Model Building and Training)

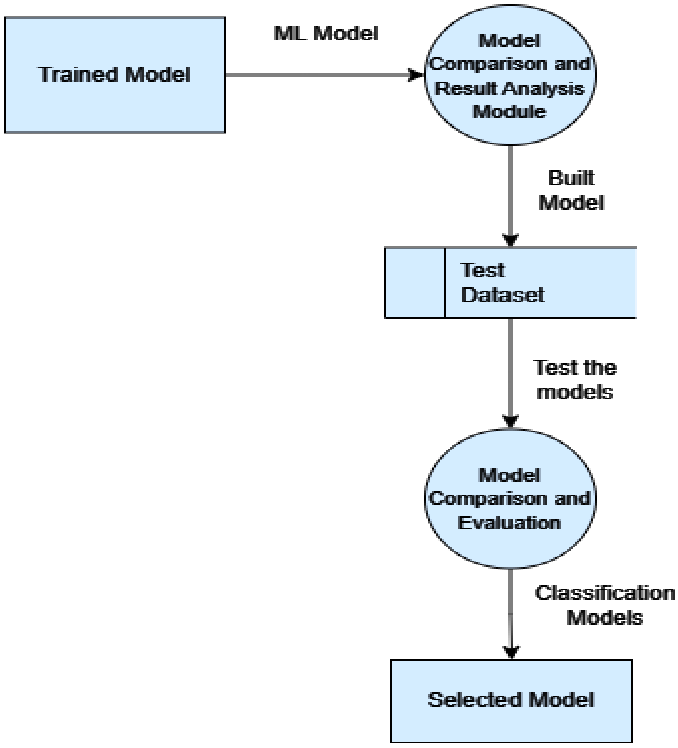

Figure 9. Level 2.4 (Model Comparison and Result Analysis)

3.5 Methodology

For the purpose of "Deciphering Crime Suspects' Emotion for Investigative Precision Using Deep Learning," the dataset containing 35,887 samples of facial expressions is divided into test, validation, and training sets in order to provide a wide range of emotion representation. In order to effectively train CNN, pre-processing converts facial expressions into standardized 48 x 48 pixel grayscale pictures and encodes emotion labels onto binary vectors.

Convolutional layers with batch normalization and activation functions are part of the CNN architecture, which is used to identify patterns linked to emotions. By concentrating on important traits, max-pooling layers lower computing complexity. Dense layers and flattening improve the model's capacity to identify intricate links.

The output layer uses suitable activation functions that are in line with the emotional subtleties of criminal interrogations to classify facial expressions into seven emotion groups. This methodology harnesses deep learning, particularly CNNs, to accurately recognize and decipher crime suspects' emotions, aiming for investigative precision.

3.6 Module Description

Login and Admin Module: This module guarantees safe entry to the system. Users authenticate themselves through a login interface, protecting private data. Admin rights make it easier to manage users, systems, and data. This module creates a strong access control system to guarantee allowed user interactions and efficient management when utilizing deep learning CNN to identify the emotions of criminal suspects.

The Data Collection Module: It is a crucial component of the research since it records a range of facial expressions that are displayed during criminal interrogations. To guarantee consent and legal compliance, ethical standards are set. A curated sample dataset with a range of nationalities and backgrounds is included. Every sample has accompanying emotional annotations, which serve as the ground truth for the Convolutional Neural Network's training and encourage the creation of objective and reliable models.

Preprocessing Module: To enable efficient CNN training, the Preprocessing Module refines gathered facial data. To standardize inputs, it makes use of methods like face identification, alignment, and normalizing. Noise and image artifacts are removed to guarantee constant quality. This module optimizes facial photos and improves data integrity, readying them for feature extraction and the project's later training phases.

Feature Extraction Module: This module finds and extracts pertinent facial features by using a Convolutional Neural Network (CNN) architecture. To do this, the network must be trained to identify subtle patterns linked to various emotions. The module improves the model's capacity to precisely interpret the emotions of criminal suspects by acquiring discriminative features, which paves the way for accurate investigation results throughout the project.

Emotion Classification Module: This module uses a Convolutional Neural Network (CNN) that has been trained to classify face emotions in real-time. It analyzes complex emotional cues through feature extraction and produces immediate predictions. This module is essential for interpreting the feelings of criminal suspects during questioning, giving detectives' quick insights for accurate investigative decision-making that leverages deep learning.

Training, Testing, and Validation Module: This module trains a Convolutional Neural Network (CNN) for emotion recognition by using the preprocessed dataset. In order to ensure model generalization, hyperparameter tuning is carried out using a validation dataset. Thorough testing evaluates the recall, accuracy, and precision of the CNN, confirming its resilience. Through this iterative process, the model is refined and performance is optimized for accurately interpreting the emotions of criminal suspects.

Reporting Module: This module creates comprehensive reports that provide an overview of the emotional exchanges that occur during questioning. It contains confidence levels, stress indicators, and timestamped emotional states. Charts and other visualizations help interpret results. Reports that are exportable and shareable give detectives a wealth of information that helps them make well-informed decisions when it comes to use deep learning CNN to accurately interpret the emotions of criminal suspects.

4. Discussion

Convolutional neural networks (CNNs) were used in this effort to develop a system for identifying the emotions of criminal suspects based on video footage. The future work of the research aims to implement the proposed design and is expected to achieve an accuracy rate of 85% in identifying basic emotions such as anger, fear, sadness, and happiness. However, accuracy differed amongst demographic groups, emphasizing the need for larger and more varied datasets in order to increase generalization and lessen bias. The system's real-time emotional assessments could improve the accuracy of investigations by giving law enforcement information about the emotional states of suspects during questioning. However, difficulties may emerge because human emotions are complex and often overlap, making it possible for the model to misread data.

5. Conclusion

The application of deep learning to interpret the emotional state of criminal suspects has the potential to improve the accuracy of law enforcement investigations. The utilization of cutting-edge technology such as artificial intelligence and machine learning has the potential to transform conventional investigative techniques and offer significant advantages in comprehending the emotional states of suspects. Deep learning models make it possible to analyze small facial clues, providing a more sophisticated understanding and maybe producing faster, more accurate results. But there are issues and concerns related to ethics that need to be addressed. It's critical to address privacy issues, algorithm bias, and the requirement for open procedures. It is critical to strike a compromise between maximizing the benefits of deep learning and defending individual rights. In conclusion, integrating deep learning to comprehend the emotions of criminal suspects can improve the efficacy of investigations. Law enforcement may use technology to better serve communities by addressing ethical problems.

References

[1] Modi, Shrey, and Mohammed Husain Bohara. "Facial emotion recognition using convolution neural network." In 2021 5th international conference on intelligent computing and control systems (ICICCS), pp. 1339-1344. IEEE, 2021.

[2] Singh, Shekhar, and Fatma Nasoz. "Facial expression recognition with convolutional neural networks." In 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), pp. 0324-0328. IEEE, 2020.

[3] Zhang, Hongli, Alireza Jolfaei, and Mamoun Alazab. "A face emotion recognition method using convolutional neural network and image edge computing." IEEE Access 7 (2019): 159081-159089.

[4] Mehendale, Ninad. "Facial emotion recognition using convolutional neural networks (FERC)." SN Applied Sciences 2.3 (2020): 446.

[5] Sang, Dinh Viet, and Nguyen Van Dat. "Facial expression recognition using deep convolutional neural networks." In 2017 9th International Conference on Knowledge and Systems Engineering (KSE), pp. 130-135. IEEE, 2017.

[6] Jaiswal, Akriti, A. Krishnama Raju, and Suman Deb. "Facial emotion detection using deep learning." In 2020 international conference for emerging technology (INCET), pp. 1-5. IEEE, 2020.

[7] Cherif, Raoudha Yahia, Abdelouahab Moussaoui, Nabila Frahta, and Mohamed Berrimi. "Effective speech emotion recognition using deep learning approaches for Algerian dialect." In 2021 International Conference of Women in Data Science at Taif University (WiDSTaif), pp. 1-6. IEEE, 2021.

[8] Begaj, Sabrina, Ali Osman Topal, and Maaruf Ali. "Emotion recognition based on facial expressions using convolutional neural network (CNN)." In 2020 International conference on computing, networking, telecommunications & engineering sciences applications (CoNTESA), pp. 58-63. IEEE, 2020.

[9] Pandey, Shubhanjay, and Sonakshi Handoo. "Facial Emotion Recognition using Deep Learning." In 2022 International Mobile and Embedded Technology Conference (MECON), pp. 248-252. IEEE, 2022.

[10] Borgalli, Rohan Appasaheb, and Sunil Surve. "Deep learning framework for facial emotion recognition using CNN architectures." In 2022 International Conference on Electronics and Renewable Systems (ICEARS), pp. 1777-1784. IEEE, 2022.

[11] Kedari, Pradnya, Mihir Kapile, Divya Kadole, and Sagar Jaikar. "Face emotion detection using deep learning." In 2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS), pp. 118-123. IEEE, 2021.

[12] Chouhayebi, Hajar, Jamal Riffi, Mohamed Adnane Mahraz, Ali Yahyaouy, and Hamid Tairi. "Facial expression recognition using machine learning." In 2021 Fifth International Conference On Intelligent Computing in Data Sciences (ICDS), pp. 1-6. IEEE, 2021.

[13] Khajuria, Oreeti, Rakesh Kumar, and Meenu Gupta. "Facial Emotion Recognition Prediction Technique: A Review." In 2023 2nd International Conference on Computational Modelling, Simulation and Optimization (ICCMSO), pp. 5-11. IEEE, 2023.

[14] Raghuvanshi, Arushi, and Vivek Choksi. "Facial expression recognition with convolutional neural networks." CS231n Course Projects 362 (2016).

[15] Fan, Yin, Xiangju Lu, Dian Li, and Yuanliu Liu. "Video-based emotion recognition using CNN-RNN and C3D hybrid networks." In Proceedings of the 18th ACM international conference on multimodal interaction, pp. 445-450. 2016.

[16] Duncan, Dan, Gautam Shine, and Chris English. "Facial emotion recognition in real time." Computer Science (2016): 1-7.

Author's biography

Harikrishnan A - B.Tech Final Year Student, Computer Science And Engineering, Jawaharlal College of Engineering and Technology (JCET) Lakkidi, Ottapalam, Palakkad.

Nikhil Suresh - B.Tech Final Year Student, Computer Science And Engineering, Jawaharlal College of Engineering and Technology (JCET) Lakkidi, Ottapalam, Palakkad.

Sanil M - B.Tech Final Year Student, Computer Science And Engineering, Jawaharlal College of Engineering and Technology (JCET) Lakkidi, Ottapalam, Palakkad.

Sumesh P- B.Tech Final Year Student, Computer Science And Engineering, Jawaharlal College of Engineering and Technology (JCET) Lakkidi, Ottapalam, Palakkad.

Remya M – Assistant Professor in Computer Science And Engineering, Jawaharlal College of Engineering and Technology (JCET) Lakkidi, Ottapalam, Palakkad.