Emotion Detection Using Deep Learning: A Review

Harsh Dhiman
Computer Science & Engineering
COER University,
Roorkee, India
hd9oct2004@gmail.com

Abhinav Garg
Computer Science & Engineering
COER University,
Roorkee, India
gargabhi291@gmail.com

Anmol Saini
Computer Science & Engineering
COER University,
Roorkee, India
anmolsaini0731@gmail.com

Ravi Kumar
Computer Science & Engineering
COER University,
Roorkee, India
rvkumar198@gmail.com

Yadika Prasad
Computer Science & Engineering
COER University,
Roorkee, India
prasadyadika@gmail.com

Kapil Kumar
Computer Science & Engineering
COER University,
Roorkee, India
kumarkapil51@gmail.com

***Abstract*— An emotion detection system, which will identify the human emotions by reading the facial expressions very accurately. The system employs sophisticated computer vision techniques and deep learning algorithms to categorize emotions such as happiness, anger, sadness, fear, and surprise in real-time video or image data. The first segment addresses face identification, the subsequent section focuses on feature extraction, and the last half executes classification using sophisticated models like Convolutional Neural Networks (CNNs) for efficient emotion recognition.**

Experiments on activity datasets show that the developed system is accurate and efficient, which is promising for applications in various fields from human-computer interaction to mental health tracking and marketing analysis. This system attempts to identify six distinct human emotions as derived from various facial expressions, voice intonations or textual data based on the input type but with an added emphasis being placed on facial expression recognition. Use cases of the field are vary, from improving customer service in call centers, through entertainment production and marketing to healthcare or mental health analysis.

***Keywords*— Histogram of Oriented Gradient (HOG), Scale Invariant Feature Transform (SIFT), Convolutional Neural Network, feature extraction.**

I. Introduction

There are several applications for human emotion recognition that call for extra protection or personal data. This may be considered an extension of face detection, necessitating the integration of an extra security layer that identifies emotions alongside facial recognition. This can be helpful in confirming that the subject of the camera is more than simply a two-dimensional image. Promotions for businesses are another significant area where emotion detection is crucial. Customers' reactions to all of their products and offerings are what most businesses rely on to succeed. An artificially intelligent system can determine if a consumer liked or hated a product or service if it can recognize and record emotions in real time from a user's picture or video [3]. As we've seen, the primary motivation for identifying someone is security. It may be based on retinal detection, voice recognition, passwords, fingerprint matching, etc. To avoid danger, it might also be crucial to determine the person's intentions.

This is beneficial in high-risk areas, such as airports, concerts, and huge public assemblies, where several breaches have lately transpired. Human emotions can be classified as: fear, contempt, disgust, anger, surprise, sorrow, happiness, and neutrality. These emotions are quite complicated.

A. Facial Emotion Recognition

Facial emotion identification is the method employed to discern human emotions using facial expressions. The human brain intrinsically perceives emotions, and recent innovations have resulted in software that can identify emotions. This technology will eventually be capable of interpreting emotions with the same efficacy as human brains with further advancements. AI can discern emotions by comprehending the significance of each facial expression and using that understanding to evaluate freshly presented information. Technology capable of interpreting, replicating, decoding, and responding to human emotions and facial expressions is referred to as emotional artificial intelligence, or emotion AI. Humans convey emotions significantly via facial expressions. Facial expression recognition is one of the most potent, instinctive, and quick methods for humans to convey their feelings and goals [4]. The recognition of facial expressions of emotion has experienced a substantial rise in study in recent years, attributable to its numerous applications in computer vision, robotics, and human-computer interaction. In a study utilizing the facial recognition technology (FERET) dataset, Paul Ekman discovered seven universal emotions—anger, fear, happiness, sorrow, neutrality, disgust, and surprise—alongside the requisite facial postures and muscular movements for these expressions (Ekman, 1997). This concept, later adopted by Ekman & Friesen (2003), functions as an efficient way for classifying human expressions. FER systems were predominantly constructed using FACS in prior applications. The Extended Cohn-Kanade dataset (CK+), the Japanese Female Facial Expressions (JAFFE) dataset, and the FER2013 dataset are significant resources for research in facial expression recognition (Canade, Cohn & Tian, 2000; Lucey et al., 2010; Goodfellow et al., 2013). Each dataset differs in terms of image kind, quantity, and tagging process. The CK+ dataset has Action Units (AUs) for each facial picture, categorizing faces in accordance with the Facial Action Coding System (FACS) methodology. The installation of the FER system poses many challenges. The bulk of datasets consist of images of posed persons exhibiting certain emotions [5]. The primary issue is that real-time applications need a model with unposed and undirected expressions. The second problem is that the dataset labels are

ISBN: 978-93-343-1044-3

broadly categorized, indicating that the system may consistently identify specific phrases in real time.

B. System Functionalities

Deep learning is a significant domain within machine learning and a crucial tool in artificial intelligence. In terms of picture identification, object detection, natural language translation, and trend prediction, deep learning excels. we employed deep learning models to extract characteristics. We also pre-processed our data using morphological operations and data augmentation techniques. This model is predicated on the convolutional neural network architecture. This system has convolutional, ReLU, and max-pooling layers. Our model comprises max-pooling layers, ReLU layers, and two sets of convolutional layers. Subsequently, we used a softmax layer, a ReLU layer, and a fully connected layer to predict the label [6]. Facial expression recognition is a process performed by humans or computers, consisting of:

1. Recognizing faces inside a scene (e.g., in an image; this process is referred to as face detection)
2. Facial feature extraction involves identifying facial component shapes and characterizing skin texture within the observed facial region.
3. Examining the dynamics of facial features and/or alterations in their appearance, and categorizing this data into various facial-expression interpretation classifications, including facial muscle activations such as smiling or Frowning, affective classifications such as happiness or wrath, and attitudinal classifications like (dis)liking or ambivalence. This procedure is often referred to as facial expression analysis.

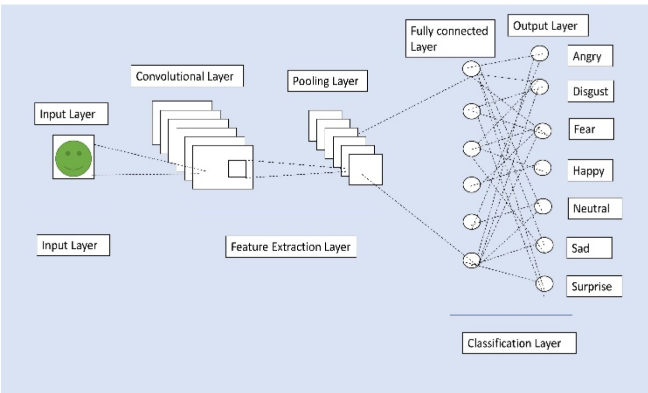


Figure 1.1Working of the System [7]

C. Convolutional Neural Networks

Neurons are the basic building blocks of a neural network. Figure 1.2 illustrates the structure of a neuron. Information propagates across a neuron when inputs are multiplied by their corresponding weights and subsequently aggregated. A bias term and a nonlinear activation function are applied to this result, shifting the output. A neuron's structure is shown in Figure 1.2. Once inputs have been multiplied by their individual weights and then merged, information moves forward via a neuron. Additionally, CNNs may approximate complicated functions by using a nonlinear function [8].

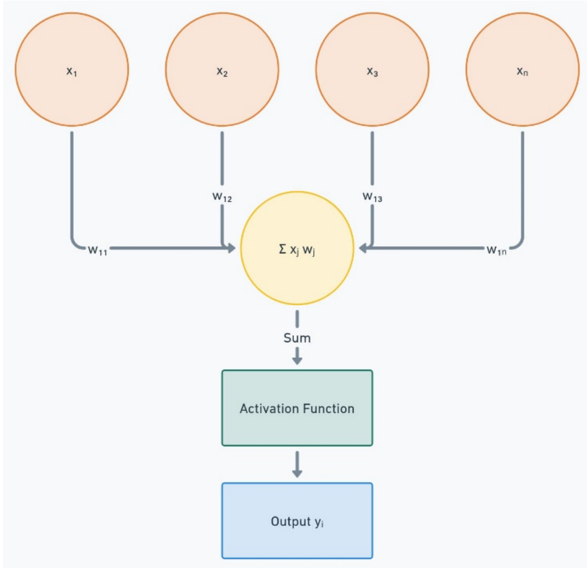


Figure 1.3 Structure of Neuron [9]

- *The Convolutional Neural Network Concept*

A CNN is a deep learning system that differentiates images by assigning varying significance to distinct features or objects within the image using learnable weights and biases. In comparison to other classification techniques, a CNN needs much less preprocessing. Figure 1.3 [8] illustrates the activities of the CNN. The architecture of the visual cortex inspired the design of the CNN, which parallels the neural connectivity in the human brain [32]. A function of a CNN is to condense images while preserving critical features necessary for precise prediction [10]. This is essential for developing an architecture capable of scaling to large datasets and proficient in feature learning. The twelve fundamental processes of CNNs include convolution, pooling, batch normalization, and dropout, which are elucidated below:

D. *CNN Architecture*

A convolutional neural network's standard architecture consists of an output layer, many convolutional layers, numerous fully connected layers, and an input layer. Several modifications were included in the LeNet Architecture throughout the development of CNN [12]. The six levels do not consider input and output. The study's Convolutional Neural Network architecture is seen in Figure 1.4.

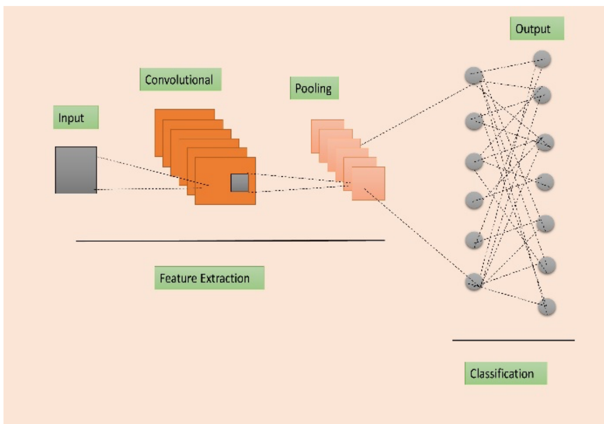


Figure 1.2 CNN Operations [11]

ISBN: 978-93-343-1044-3

- Input Layer: The image requires pre-processing before to being entered into the layer due to its predetermined dimensions. Normalized grayscale 48 × 48 pixel pictures from the Kaggle dataset are used for training, validation, and testing. The OpenCV Haar Cascade Classifier is used to recognize, crop, and normalize faces in the test photos taken by the suggested laptop webcam.

- Convolution and Pooling (ConvPool) Layers: Convolution and pooling are done in batches. There are N pictures in each batch, and the CNN filter weights are changed for each batch. The four dimensions of the image batch input received by each convolution layer are N, Color-Channel, width, and height. The feature maps or filters used in convolution are four-dimensional, as are the number of input feature maps, output feature maps, filter width, and filter height. Each convolutional layer performs four-dimensional convolutions between feature maps and image batches. Only the image's width and height are altered after convolution.

New picture width = old picture width - filter width + 1
New picture height = previous picture height - filter height + 1

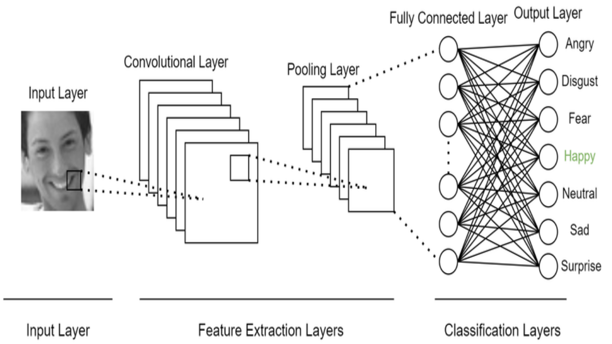


Figure 1.4 CNN Architecture [13]

Down sampling and subsampling are used after each convolutional layer to decrease dimensionality. Pooling is the name given to this technique. The two most common pooling strategies are average pooling and maximum pooling. This project uses max pooling after convolution. When the pool size is set to (2 × 2), the image is divided into a grid of blocks, each of which may have a maximum size of 4 pixels. Post-pooling, only width and height are impacted. A pooling layer and two convolutional layers are used in the architecture. N x 1 x 48 x 48 are the starting convolution layer dimensions of the batch of input images. The image's batch size is N, its width and height are 48 pixels, and it has one color channel. A batch of images with dimensions of N x 20 x 44 x 44 is produced by convolution with a feature map of dimensions 1 x 20 x 5 x 5. Batch measurements of the resulting image are (N x 20 x 22 x 22) after convolution pooling with a pool size of (2 x 2). A feature map with dimensions of 20x20x5x5 is created by applying an additional convolution layer, yielding an image batch of size (N x 20 x 18 x 18). The resulting picture batch has the size (N x 20 x 9 x 9) after a pooling layer with a pool size of (2 x 2).

- Fully Connected Layer: The mechanism by which neurons transmit signals throughout the brain served as the prototype for this layer. It employs layers interconnected by adjustable weights to transform a multitude of input features. Two hidden layers—500 and 300 units, respectively—are used by fully connected layers. The weights of these layers are adjusted via forward propagation of training data and backward propagation of mistakes. Backpropagation initially evaluates the difference between the actual value and the prediction in order to determine the appropriate weight modifications for each previous layer. The learning rate, momentum, regularization parameter, and decay are the hyperparameters for this layer. The output of the second pooled layer is Nx20x9x9, whereas the input of the first hidden layer of the fully-connected layer is Nx500. Thus, the output from the pooling layer is flattened to Nx1620 size and then sent to the first hidden layer. The first hidden layer provides data to the second hidden layer. The output layer receives the second hidden layer's output, which has dimensions of Nx300 and represents the number of facial expression classes.

- Output Layer: There are seven distinct classes in the output layer, which is connected to the output from the second hidden layer. The probability given to each of the seven classes is used by the Softmax activation function to produce output. The class that is most likely to occur is the one that is expected.

E. *Library And Packages*

- OpenCV: OpenCV (Open Source Computer Vision Library) is an open-source software library designed to establish a unified infrastructure for computer vision applications and to accelerate the implementation of machine perception in commercial products. It includes more than 2500 optimized algorithms, including a wide range of both conventional and state-of-the-art machine learning and computer vision methods. Due to its BSD license, OpenCV is easily accessible for enterprises to use and alter the source. These algorithms facilitate a variety of tasks, including face detection and identification, object recognition, human action classification in videos, tracking camera movements and moving objects, 3D model extraction, 3D point cloud generation from stereo cameras, image stitching for high-resolution scene creation, database retrieval of similar images, red-eye removal from flash photography, eye movement tracking, and the creation of markers for augmented reality overlays. OpenCV employs MMX and SSE instructions when accessible and is mostly designed for real-time vision applications [14]. Efforts are now being made to build a fully

ISBN: 978-93-343-1044-3

operational CUDA and OpenCL interface. Applications of OpenCV encompass:
- Two-dimensional and three-dimensional feature toolkits
- Facial recognition system
- Gesture recognition
- Human-computer interface (HCI)
- Object recognition
- Motion tracking
- Augmented reality

OpenCV includes a statistical machine learning package that encompasses several of the previously stated domains. It comprises: -
- Decision tree learning
- k-nearest neighbor method
- Naive Bayes classifier
- Artificial neural networks
- Stochastic forest Random Forest
- Support Vector Machine (SVM)
- Deep Neural Networks (DNN)

- Numpy: NumPy stands for "Numeric Python" or "Numerical Python." Rapid precompiled functions for mathematical and numerical calculations are offered by this open-source Python extension package. Furthermore, NumPy introduces strong data structures to Python, augmenting the programming language to enable efficient computation of multi-dimensional arrays and matrices [15]. The method also targets large arrays and matrices. It is the core module for Python scientific computing. Among its many aspects are the following significant ones:
  - A highly potent N-dimensional array object with advanced broadcasting capabilities
  - Instrumentation for combining Fortran and C/C++ code Strong skills in linear algebra, stochastic processes, and the Fourier Transform

- Keras: Based on Python Keras is a high-level neural network API compatible with TensorFlow, CNTK, or Theano. The objective of its creation was to enable fast experimentation. A number of commonly used elements for building neural networks are implemented by Keras, including layers, goals, optimizers, activation functions, and a variety of tools for processing text and picture input. A Slack channel and the GitHub problems page are two community support channels for the code, which is maintained on GitHub. Users may productize deep models utilizing Keras on web-based platforms, Java virtual machines, and smartphones (iOS and Android).

- SciPy: NumPy and SciPy (Scientific Python) are frequently used interchangeably. With additional helpful functions for regression, minimization, Fourier transformation, and many more areas, SciPy expands on NumPy's capabilities. NumPy is derived from two previous array-related Python libraries. Numerical is one such. Although it is no longer in use, Numeric is similar to NumPy, a Python package for high-performance numerical computing. Numarray, a deprecated version of Numeric that is a complete rewriting of Numeric, is another antecedent to NumPy.

- Tensor Flow: Google created and released TensorFlow, a Python library for rapid numerical computing. This foundational library, constructed atop TensorFlow, may be utilized for the direct development of Deep Learning models or to facilitate the process using wrapper libraries.

- Haar Cascade Classifier in OpenCv: In the effective machine learning method known as Haar feature-based cascade classifiers, a cascade function is learned from a sample that includes a significant number of both positive and negative pictures. Cascade classifiers are generated as a consequence of AdaBoost classifiers. by segmenting the strong classifiers into phases. This kind of classifier is called "cascade" because it consists of a number of smaller classifiers that are applied to the area of interest until the selected item is either rejected or passed. Training and detection are the two phases into which the cascade classifier divides the classification task. The task of collecting samples that fall into the positive and negative categories is completed by the training stage [17]. The cascade classifier makes use of a few supporting functions to generate a training dataset and evaluate the predominance of classifiers. A collection of both positive and negative samples is required for training the cascade classifier.

II. REVIEW OF LITERATURE

The literature review contextualizes a dissertation by finding previous research and elucidating the study of a certain topic. In a mixed methods study, the literature review is characterized as a consistent methodology applicable to either quantitative or qualitative research. Comprehensive literature research was conducted utilizing credible sources, including SCI, UGC, IEEE publications, conference papers, and books focused on video security and summarization, to gather pertinent material [18]. This literature assessment offers a comprehensive overview of the current research in the domains examined in the conducted experiments arranged in Table 1.

Convolutional Neural Networks (CNNs) are often the chosen neural network design over alternatives like ResNet, MobileNet, or attention-based models. Their allure is found in the way they combine functionality and performance to meet certain needs.

CNNs provide a simple structure with fully connected layers to wrap it up, pooling to simplify, and convolutional layers to identify patterns. Unlike MobileNet's lightweight architecture, which is optimised for devices with limited resources, or ResNet's deeper, skip-connection-driven

ISBN: 978-93-343-1044-3

Table 1 Research Paper Reviews

S.No.	Author	Methodology	Advantage
1.	*Akriti Jaiswal et al.[2020]*	AI, FER, CNN, ReLu.	Model is producing state-of-art-effects on both two datasets.
2.	*Prince Awuah Baffour1 et al.[2022]*	Use CNN-based deep learning models, transfer learning, and multimodal emotional cues for robust facial emotion recognition.	Accurate FER through CNNs, multimodal cues, and transfer learning.
3.	*Jiawen Deng1 et al.[2020]*	MC-ESFE for feature extraction and ECorL for emotion correlation learning.	Enhanced multi-label emotion identification by the utilization of emotion-specific characteristics and correlation learning.
4.	*Alaa Alslaity & Rita Orji. [2022]*	Systematic review of 123 papers analyzing trends, techniques, datasets, and metrics.	Comprehensive review of ML-based emotion detection trends, methods, and challenges.
5.	*Leila Ismail et al.[2023]*	High accuracy provides population-level emotion insights, aids real-time policymaking.	High accuracy provides population-level emotion insights, aids real-time policymaking.
6.	*Md. Forhad Ali et al.[2022]*	Employs a Deep Learning model including a Convolutional Neural Network (CNN) to identify emotions by re-training on authentic face data, enhancing detection precision via effective data mining and instantaneous feedback.	enhances emotion recognition accuracy and efficiency by leveraging real-time feedback
7.	*Maliha Asad et al. [2022]*	To train and anticipate the seven main emotions—disgust, anger, contempt, fear, happiness, sorrow, and surprise—use the one versus all (OVA) method.	Improves facial emotion recognition accuracy by using one-versus-all (OVA) approach and the Viola-Jones framework
8.	*Kangning Yang et al.[2021]*	Applying various realistic distortions (e.g., noise, blur, occlusion, changes in brightness, and contrast) to facial image datasets.	Provides a comprehensive evaluation of the robustness of commercial emotion detection systems,

design, which performs well in complicated scenarios, their simplicity makes them efficient and approachable. CNNs provide results without needless overhead for projects with low computational demands.

III. METHODOLOGY

A. Datasets

Large volumes of training data are necessary for neural networks, especially deep networks. Moreover, a substantial aspect of the final model's efficacy is contingent upon the choice of images utilized for training. It indicates that a high-quality, quantitative data set is required. Research on emotion recognition may be conducted using a variety of datasets, from tens of thousands of smaller pictures to a few hundred high-resolution photographs.

The Feb2013 dataset was obtained through the Facial Expression Recognition Challenge, a Kaggle recognition contest. The collection consists of grayscale photos of female faces at different ages. The resolution of each image is 48 × 48 pixels. This data is categorized into seven subcategories: anger, disgust, fear, happiness, sorrow, surprise, and neutral. Out of 32,284 photographs in February 2013, 28,702 are allocated for validation, whereas 3,582 are designated for testing. The collection contains faces in every orientation and at every age. Feb2013 has excellent facial expression recognition capabilities.

A. System Design

The system's overall architecture is shown in the system design. The design features of the system are explained in more detail in this section. There are 34,488 photographs in the training dataset and 1,250 in the testing dataset. Happiness, sorrow, anger, surprise, neutrality, contempt, and fear are the categories into which the emotions fall. Various facial features, including elevated cheeks, nasal wrinkles, widely opened eyes, expanded lips, and arched eyebrows, are employed to express distinct emotions. Extensive datasets were utilized to enhance precision and provide the object classification for a given image. It uses max pooling and convolution layers based on those features [19]. These seven universal emotions are represented by the following expressions in Fig. 3.2.

B. Emotion Detection Using Deep Learning

This research employs a powerful CNN for image identification with Google's "Keras" deep learning open-source toolkit for face emotion recognition. We trained our proposed network on two separate datasets and evaluated its loss and validation accuracy. We employed an emotion model created by a convolutional neural network (CNN) using deep learning to detect facial expressions in photos sourced from a dataset encompassing seven distinct emotions [20]. Utilizing the Keras library offered by Google, we have modified the CNN architecture to enhance accuracy and adjusted many steps in the CNN process. With the suggested network, we used Keras to implement emotion detection.

C. Steps To Design Emotions Detection System Data Collection

Collect or choose an appropriate dataset for training the emotion detection model. Popular datasets include FER2013, CK+, and AffectNet. These datasets must encompass a variety of face expressions linked to distinct emotions. Perform data augmentation to increase data diversity, such as by rotating, flipping, or changing the brightness of images.

ISBN: 978-93-343-1044-3

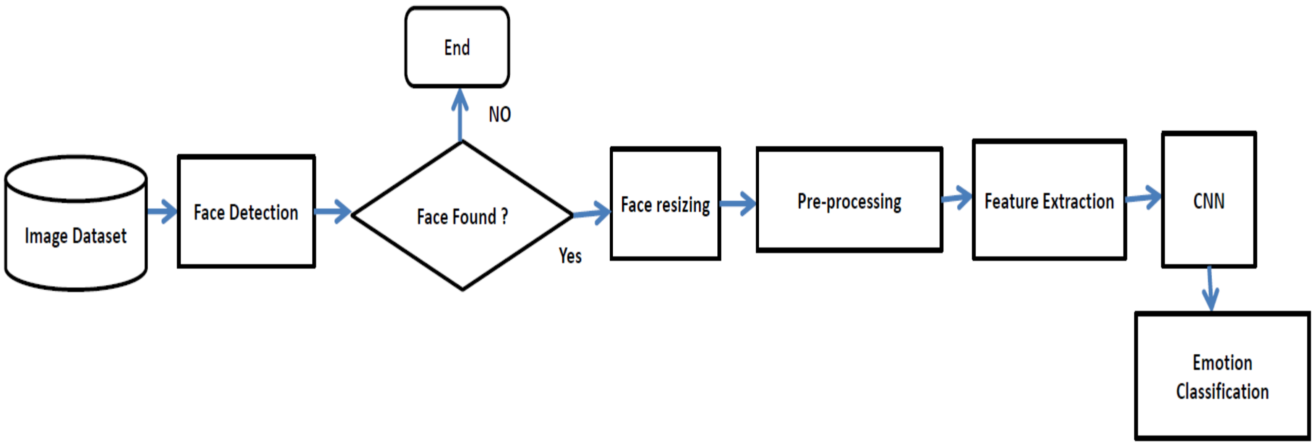


Figure 3.2 Emotion Detection System Design

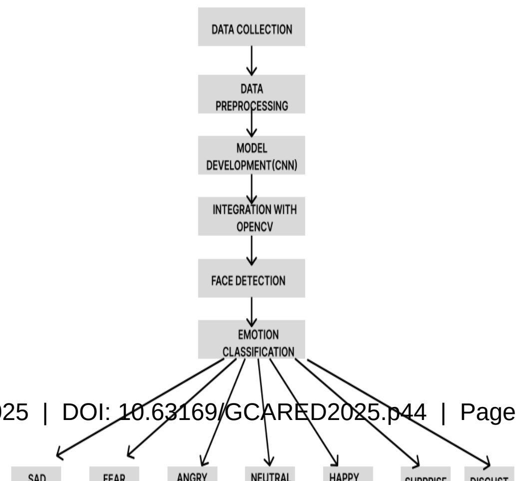
i. Data Preprocessing
   - Resize images to a consistent input size (e.g., 48x48 pixels).
   - Convert photos to grayscale when utilizing a model that does not require color information.
   - Normalize pixel values to a range of [0, 1] or [-1, 1] to enhance model training.
   - Detect and crop faces in the images using OpenCV's face detection algorithms (e.g., Haar Cascades or DNN-based face detection) to focus on facial regions.

ii. Model Development
   - OpenCV-based Approach: Use feature extraction methods like Local Binary Patterns (LBP) or Histogram of Oriented Gradients (HOG) to identify key facial features. To classify emotions, use a classifier on the retrieved attributes, such as Support Vector Machine (SVM) or k-Nearest Neighbors (k-NN).

   - CNN-based Approach: Create a model for a Convolutional Neural Network (CNN) including classification and feature extraction layers. Activation functions such as ReLU are often used in convolutional, pooling, and fully connected layers.

iii. Model Architecture

Give specifics on the model architecture, including the number of layers, activation functions, regularization dropout, optimizer (like Adam), and loss function (such categorical cross-entropy).



iv. Model Training
   - Divide the dataset into training, validation, and testing subsets (e.g., 70% for training, 15% for validation, 15% for testing).
   - Train the model for many epochs using a specified batch size, while observing training and validation accuracy and loss to mitigate overfitting.
   - Employ strategies such as early halting or learning rate adjustment contingent upon validation loss to enhance training optimization.

v. Model Evaluation
   - Assess the model with measures such as accuracy, precision, recall, and F1-score.
   - Construct a confusion matrix to illustrate performance among several emotion categories and pinpoint misclassifications.
   - Calculate the mean accuracy on the test set to assess generalization performance.

vi. Real-time Implementation
   - Integrate the trained model with a webcam or video feed to enable real-time emotion detection.
   - Use OpenCV to capture live frames, preprocess each frame, and pass it through the model to predict emotions.
   - Display predicted emotion labels and bounding boxes around detected faces on the live video feed.

vii. Results and Analysis
   - Analyze the model's performance by comparing evaluation metrics and visualizing examples of correct and incorrect predictions.
   - Discuss any observed patterns, limitations, or sources of error, such as difficulty in detecting subtle emotions.

IV. RESULTS AND DISCUSSION

The measures used to assess the model's performance are described in this chapter. The training results are then used to identify each model's optimal parameter values. The CNN model's accuracy and loss are assessed using these metrics.

ISBN: 978-93-343-1044-3

The outcomes of various models are then contrasted and examined.

a. Evaluation Metrics

The metrics employed to evaluate the model's performance encompass accuracy, loss, precision, recall, and F-score. The following is a definition of these measurements [21].

Accuracy: Accuracy is given by

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \qquad (1)$$

Loss: The loss function, categorical cross-entropy, is supplied by

$$Loss = -\frac{1}{N}\sum_{i=1}^{N} y_i log/\square\hat{y}_i + (1-y_i) log\,(1-\hat{y}_i) \qquad (2)$$

Let m represent the number of classes (happy, sad, neutral, frightened, and furious), p denote the anticipated probability, and y signify a binary indicator (0 or 1).

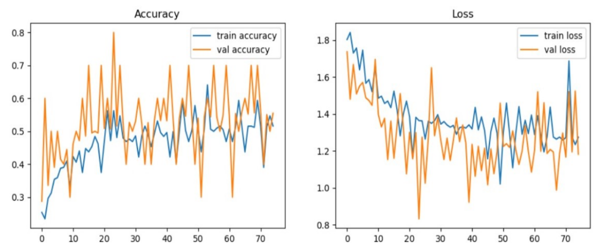


Figure 4.1 Evaluate the accuracy and loss

b. Confusion Matrix

The four sets of actual and expected values—True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN)—are represented by the confusion matrix. True positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) are used to compute precision, recall, and F-score. Whereas TN shows the correct identification of a fake emotion, FN indicates the incorrect identification of a false emotion, TP indicates the accurate prediction of an emotion, and FP indicates the incorrect prediction. Look at a picture of the lively class. Figure 4.1 illustrates the confusion matrix of the situation. Given that the cheerful image is anticipated to evoke happiness, the TP value is in the red segment. The blue segment possesses FP values because to the anticipated feelings of grief, anger, neutrality, or terror conveyed by the picture. As the image is neither gloomy, angry, neutral, nor terrifying, the yellow 22 area possesses TN values, as predicted by the model [22]. The picture, intended to be joyful, is not, as the green section contains FN values.

c. Accuracy

Processed 35,500+ facial images representing six primary emotions to train the model from Convolutional Neural Network(CNN), achieving **80% Accuracy** on test data.

d. Research Gap

Most emotion detection models are trained on datasets that lack diversity in age, gender, and culture, making it difficult to generalize in real-world scenarios. More inclusive datasets and bias-mitigation techniques are needed.

Current systems rely mainly on facial expressions, ignoring speech, body language, and physiological signals.

Synchronizing multiple data sources remains a challenge, requiring improved fusion techniques for better accuracy.

e. Disadvantage

i. Most emotion detection models are trained on limited datasets, making them less accurate for diverse users.

ii. These systems mainly focus on facial expressions and ignore speech or body language, reducing accuracy in real-world scenarios.

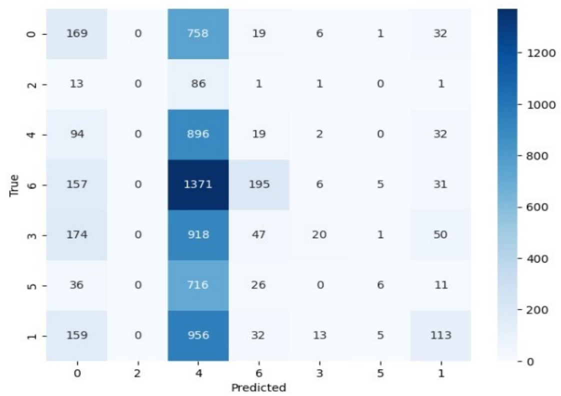


Figure 4.2 Confusion Matrix of the system

REFERENCES

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

[2] Baffour, Prince Awuah, et al. "A survey on deep learning algorithms in facial Emotion Detection and Recognition." Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi 7.1 (2022): 24-32.

[3] Alslaity, Alaa, and Rita Orji. "Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions." Behaviour & Information Technology 43.1 (2024): 139-164.

[4] Ali, Md Forhad, Mehenag Khatun, and Nakib Aman Turzo. "Facial emotion detection using neural network." the international journal of scientific and engineering research (2020).

[5] Shehata, Mohamed. "Minorities' GenderStereotypes In Media:–A natural language processing (NLP) approach to Understanding the Middle Eastern Minorities' gender stereotypes in Western Media." (2024).

[6] Oh, SeungJun, and Dong-Keun Kim. "Comparative analysis of emotion classification based on facial expression and physiological signals using deep learning." Applied Sciences 12.3 (2022): 1286.

[7] Cîrneanu, Andrada-Livia, Dan Popescu, and Dragoș Iordache. "New trends in emotion recognition using image analysis by neural networks, a systematic review." Sensors 23.16 (2023): 7092.

[8] Pal, Shantanu, Subhas Mukhopadhyay, and Nagender Suryadevara. "Development and progress in sensors and technologies for human emotion recognition." *Sensors* 21.16 (2021): 5554

[9] Shahzad, H. M., et al. "Hybrid facial emotion recognition using CNN-based features." *Applied Sciences* 13.9 (2023): 5572.

[10] Płaza, Mirosław, et al. "Machine learning algorithms for detection and classifications of emotions in contact center applications." *Sensors* 22.14 (2022): 5311.

ISBN: 978-93-343-1044-3

[11] Kola, Durga Ganga Rao, and Srinivas Kumar Samayamantula. "A novel approach for facial expression recognition using local binary pattern with adaptive window." *Multimedia Tools and Applications* 80.2 (2021): 2243-2262.

[12] Said, Yahia, and Mohammad Barr. "Human emotion recognition based on facial expressions via deep learning on high-resolution images." *Multimedia Tools and Applications* 80.16 (2021): 25241-25253.

[13] Wagle, Shivali Amit, et al. "Classification of plant leaves using new compact convolutional neural network models." *Plants* 11.1 (2021): 24.

[14] Qazi, Awais Salman, et al. "Emotion detection using facial expression involving occlusions and tilt." *Applied Sciences* 12.22 (2022): 11797.

[15] Saeed, Saad, et al. "Automated facial expression recognition framework using deep learning." *Journal of Healthcare Engineering* 2022.1 (2022): 5707930.

[16] Arul Vinayakam Rajasimman, Mayuri, et al. "Robust facial expression recognition using an evolutionary algorithm with a deep learning model." *Applied Sciences* 13.1 (2022): 468.

[17] Alonazi, Mohammed, et al. "Automated Facial Emotion Recognition Using the Pelican Optimization Algorithm with a Deep Convolutional Neural Network." *Electronics* 12.22 (2023): 4608.

[18] Haq, H. B. U., et al. "Enhanced real-time facial expression recognition using deep learning." *Acadlore Trans. Mach. Learn* 3.1 (2024): 24-35.

[19] Naseem, Muhammad Tahir, Chan-Su Lee, and Na-Hyun Kim. "Facial Expression Recognition Using Visible, IR, and MSX Images by Early and Late Fusion of Deep Learning Models." *IEEE Access* (2024).

[20] Sowmiya, R., G. Sivakamasundari, and V. Archana. "Facial emotion recognition using deep learning approach." *2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)*. IEEE, 2022.

[21] Juyal, Prachi, and Amit Kundalya. "Emotion Detection from Text: Classification and Prediction of Moods in Real-Time Streaming Text." *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2023.

[22] Kumari, Shalu, Nitika Kapoor, and Ravinder Saini. "Emotion Recognition from Facial Expression Using Deep Learning Model: A Review." *2024 5th International Conference for Emerging Technology (INCET)*. IEEE, 2024

[23] Sharma, Nittin. "An Analysis in the Identification of Secondary RNA Structure Using Energy Algorithm." *2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)*. Vol. 5. IEEE, 2024.

[24] Meena, Gaurav, et al. "Identifying emotions from facial expressions using a deep convolutional neural network-based approach." *Multimedia Tools and Applications* 83.6 (2024): 15711-15732.

[25] Prasad, S. Babu Rajendra, and B. Sai Chandana. "Mobilenetv3: a deep learning technique for human face expressions identification." *International journal of information technology* 15.6 (2023): 3229-3243.

ISBN: 978-93-343-1044-3