International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

DEEPFER: Camera Vision Based Automatic Facial Expression Analysis System Using Deep Learning

Mrs. K. Deepika[1], Mr. P. Anbumani[2]

[1] (M.C.A), [2]M.C.A., M.Phil,NET., Associate Professor.
Department of MCA, Krishnasamy College of Engineering and Technology, Cuddalore.

ABSTRACT

For human wellbeing, emotional care is crucial at every age. In a number of industries, including healthcare, human-computer interaction (HCI), security, and defence, automatic emotion recognition is essential. For facial expression recognition (FER), it is crucial and difficult to capture the dynamics of facial emotion advancement in video. In disciplines like psychology, human-computer interaction, and affective computing, facial expression analysis has attracted a lot of interest. The constructed features and rule-based algorithms that were frequently used in traditional facial expression recognition techniques had drawbacks when it came to effectively capturing the minute variations in human expressions. However, with the introduction of deep learning methodologies, particularly CNN, substantial advancements have been made in the pursuit of more precise and reliable facial emotion identification.

The Deep FER system attempts to automate and improve the interpretation of facial expressions using advancements in deep learning and computer vision, opening up a wide range of applications in numerous fields. In this research, a brand-new, low-cost, and multi-user framework for emotion care is proposed, built on big data analysis for patient feelings and based on the facial expressions that show emotion. The system uses deep learning techniques on emotional big data to extract emotional traits and recognise six different facial expressions in real-time and offline, including those expressing anger, contempt, fear, happiness, sadness, surprise, and neutrality. To train the DCNN model, a fresh dataset for emotion recognition is gathered.

The Deep FER system is an innovative web cam-based automatic facial expression analysis system that employs Convolutional Neural Networks (CNN) to accurately recognize and analyse facial expressions in real-time. The entire face observation is then applied to a deep convolutional neural network (DCNN) to learn the general properties of six different express

Keywords: Emotion AI , human computer interaction , Face Expression Recognition

1. INTRODUCTION

Affective computing is the study and development of systems and devices that can recognize, interpret, process, and simulate human affects. It is an interdisciplinary field spanning computer science, psychology, and cognitive science[1]. While the origins of the field may be traced as far back as to early philosophical inquiries into emotion ("affect" is, basically, a synonym for "emotion."), the more modern branch of computer science originated with Rosalind Picard's 1995 paper on affective computing[2].A motivation for the research is the ability to simulate empathy. The machine should interpret the emotional state of humans and adapt its behavior to them, giving an appropriate response for those emotions.[3]

Affective computing technologies sense the emotional state of a user (via sensors, microphone, cameras and/or software logic) and respond by performing specific, predefined product/service features,such as changing a quiz or recommending a set of videos to fit the mood of the learner.

Affective computing is human-computer interaction in which a device has the ability to detect and appropriately respond to its user's emotions and other stimuli[4].A computing device with this capacity could gather clues to user emotion from a variety of sources[5].

Facial expressions, posture, gestures, speech, the force or rhythm of key strokes and the temperature changes of the hand on a mouse can all signify changes in the user's emotional state, and these can all be detected and interpreted by a computer. A built-in camera captures images of the user and algorithms are used to process the data to yield meaningful information. Speech recognition and gesture recognition are among the other technologies being explored for affective computing applications.[6]

Recognizing emotional information requires the extraction of meaningful patterns from the gathered data[7]. This is done using machine learning techniques that process different modalities, such as speech recognition, natural language processing, or facial expression detection.

1.1 Industry-Specific Use Cases

As emotion AI technology can be introduced to many different industries with a variety of applications, tech giants and startups have started to invest in either computer vision or voice analysis to recognize human emotions. As a result, the technology has grown rapidly in the past two years and expanded into various new areas and industries to help businesses offer better customer experience and achieve real cost savings.

1.1.1 Customer Service

- Intelligent call routing: Businesses can detect angry customers from the beginning of the call, and such calls can be routed to more experienced and well-trained call agents.
- Recommendations during calls: Emotion AI can also provide suggestions about handling customer calls based on similar speech patterns during the conversation.
- Continuous improvement: Reviews are time-consuming and completed by only a small share of customers. Amazon sellers share that only around 3-5% of their buyers leave product reviews. Like analyzing written reviews, emotion AI can also measure how effective the calls are and if the customer is satisfied at the end of the call by leveraging voice analysis. This data can be used to improve customer services even in cases where customers do not leave reviews.

1.1.2 Human Resources

- **Recruitment:** Businesses can observe how stressful candidates are and how they communicate emotions during interviews to make better recruitment decisions. Unilever is one of the companies that is currently using emotion AI during job interviews.
- **Employee training:** Affective computing can be used for training employees who will interact directly with customers. Employees work with intelligent customer interaction simulations that evolve based on the employees' responses and emotions, helping them improve their empathy and customer service skills.
- **Tracking employee satisfaction:** HR teams can track employees' stress and anxiety levels during the job and observe if they are satisfied with their current tasks and workload. However, it also brings an ethical issue of monitoring all employees during work hours and might require their consent to monitor their emotions continuously.

1.1.3 Healthcare

- **Patient care:** A bot can be used not only for reminding patients to take their medications but also to monitor their physical and emotional well-being every day to observe if there are any problematic issues.
- **Medical diagnosis:** Affective computing can leverage voice analysis to help doctors diagnose diseases like depression and dementia.
- **Counseling:** Emotion AI can be used in counseling sessions to track and understand mental states better and help doctors support counselee more effectively.

1.1.4 Autonomous driving / Driver assistance

- Safety: Automotive companies can leverage computer vision to track the driver's emotional state while driving. If the driver is too tired, stressed, or angry/sad, it can provide alerts for unsafe driving.
- Driving performance: Affective computing can also be used for measuring the driving performance of autonomous cars. With cameras and microphones embedded in the vehicle, the technology can monitor the passengers' emotional state and observe if they seem stressed or satisfied with the driving experience.

1.1.4 Education

- Measuring effectiveness: Sensors like video cameras or microphones can be used for students' emotional states during lessons. Emotion AI can assess how satisfied or frustrated students are with the lessons because a task is too challenging or too simple. As a result, teachers can adapt themselves to tailor class load accordingly. A similar approach can also be used while testing learning software prototypes for online learning.
- Supporting autistic children: Another use case in education is to help autistic children recognize other people's emotions in the school environment.

2. PROBLEM STATEMENT

Human beings are mostly emotional, and our social interaction is measured by taking into consideration our ability to communicate emotions and to perceive the emotional states of others. Affective computing provides computing systems with mechanisms that emulate and/or interpret human emotions[1]. Its main objective is to make communication with computing systems easier and more natural. Combination, starting from facial expressions, oral intonation, psycho-physiological information, or even the texts used. To make them known to users, it is usual to employ avatars and speech synthesis, frequently combining the two. Although, in general, people are experts in recognizing and expressing emotions, sometimes there is misunderstanding when transmitting them. [2]This may be caused by ambient issues (noise, lighting, or distance between interlocutors), or even personal issues (concentration or the behavior or confidence with the interlocutor). This is why emotional resources are frequently validated by people, in order to ascertain whether they really express the correct emotion or if the interlocutors are able to perceive them adequately[3]. Many times, resources are not very expressive or not correctly understood by humans; therefore, computing Facial expression is one of the most natural and immediate means for human beings to communicate their emotions, as the human face can express emotions sooner than people verbalize or even realize their feelings.

Automatic facial expression recognition (FER) has become an increasingly important research area that involves computer vision, machine learning, and behavioral sciences. Much progress has been made in building computer systems to understand and use this natural form of human communication, although most of these systems attempt to recognize only a small set of prototypical emotional expressions[4]. FER can be used for many applications, such as security, human-computer interaction, driver safety, and health care.

Emotions are multimodal, and currently, they are recognized within the Human-Computer Interaction (HCI)area using the following factors, both separately and in emotion plays an important role in human life and communication[5]. In the daily life, emotion is an inextricable part of the interaction of human beings, which can be observed by the changes in physiological features and behaviors. Because emotion recognition has a great potential to improve the quality of our life, in the past decades, emotion recognition has aroused a lot of attention of many researchers and has been a popular research topic in various fields such as robotics, human-computer interaction, and entertainment, to name a few[6]. Meanwhile, emotion care can be very useful in medical applications when medical staff need to assess the patient's feeling and behavior during or after the surgery[7].With the development of big data and deep learning, huge amount of data including emotional data is generated in recent years, which cannot be handled with the traditional techniques.

3. PROPOSED METHODOLOGY

This section presents the proposed deeply learned classifiers for facial emotion classification of unfiltered real-life face images.

The author propose a model that uses DCNN architecture to predict the emotion of human's faces from unfiltered real-world environments. The novel CNN approach addresses the Emotion labels as a set of discrete annotations and train the classifiers that predict the human's Expressions.The author design a quality and robust image pre-processing algorithm that prepare and pre-process the unfiltered images for the CNN model and this greatly has a very strong impact on the performance accuracy of our facial Emotion classifiers. The proposed method uses DeepCNN to build a FER scheme. The presented model can be used in real-time using a webcam to categorize human faces.

- Region Proposal Network is used to detect the face.
- Local Binary Pattern is used to extract of face and emotion feature.
- DCNN is used to classify the face and emotions.
- Recommendation System using predicted facial expressions.

Advantages

- The detection speed and accuracy have been greatly improved.
- Greatly reduced the computational complexity.
- Accurate predictions.
- Have flexibility and robustness, together with efficiency in inference time.

4. MODULES

4.1 Facial Expression Dashboard

This module describes how a web user interface for a dashboard that displays facial expression often comprises of a number of modules that operate in concert to give consumers an intuitive and simple to use interface.

4.1.1 Dashboard module: The dashboard module is the main interface for the facial expression recognition system, and it displays the results of the facial expression analysis in real-time.

4.1.2 Database module: The database module stores the data related to the facial expressions analyzed, such as the timestamp, facial expression, and other relevant details.

4.1.3 User interface module: The user interface module provides a graphical user interface for the dashboard, allowing users to interact with the system and access the results of the facial expression analysis.

4.2 Facial Emotion Classification -Training Phase

4.2.1 Training Dataset

Facial Expression Recognition 2013 (FER-2013) dataset was prepared in Challenges in Representation Learning: Facial Expression Recognition Challenge, which is hosted categories (e.g., angry, disgust, fear, happy, sad, surprise, and neutral) and three different sets such as training set (28.709 images), validation set (3.589 images), and test set (3.589 images).

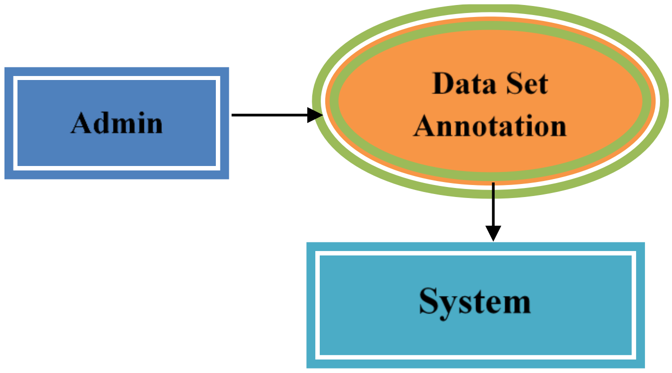

Fig. No.: 1 Diagrammatic representation of Training Dataset

All images in this dataset are grayscale with 48 X 48 pixels, thus corresponding to faces with various poses and illumination, where several faces are covered by hand, hair, and scarves[1]. Because of FER-2013 is collected from the Internet and has various real-world conditions, it becomes one of the largest and most challenging databases for facial expression recognition.

4.2.2 Pre-Processing

Image preprocessing are the steps taken to format images before they are used by model training and inference. The steps to be taken are,

- Read image
- RGB to Grey Scale conversion
- Resize image Original size (360, 480, 3) — (width, height, no. RGB channels) Resized (220, 220, 3)
- Remove noise(Denoise) -smooth our image to remove unwanted noise using gaussian blur.
- Binarization Image binarization is the process of taking a grayscale image and converting it to black-and-white, essentially reducing the information contained within the image from 256 shades of gray to 2: black and white, a binary image.[2]

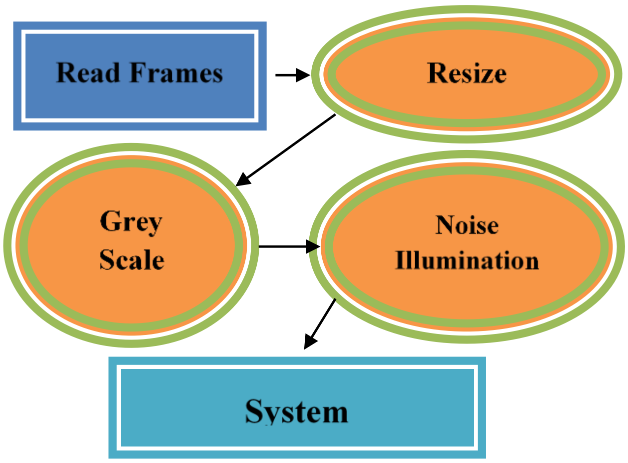

Fig. No.: 2 Diagrammatic Representation of Pre-Processing

- In that will enhance the different features of images and get for example its intensity, contrast, saturation for different image processing[3]. Low pass-filters a grayscale image that has been degraded by constant power additive noise and uses a pixel wise adaptive Wiener method based on statistics estimated from a local neighborhood of each pixel.[4]

4.2.3 Face Expression and Detection

The Background subtraction approach is mostly used when the background is static. The principle of this method is to use a model of the background and compare the current image with a reference[4]. The foreground objects present in the scene are detected and attempts to detect moving regions in an image by differencing between current image and a reference background image in a pixel-by-pixel fashion[5]. And use the static background for the image subtraction which will give the human have to track.

This step detects objects of interest as they move about the scene[6]. The action detection process is independently applied to all the static cameras present in the scene. For human recognition is feature extraction and representation where the important characteristics of image frames are extracted and represented in a systematically way as features[7].First, an efficient multitask region proposal network (RPN), combined with boosting face detection, is developed to obtain the human face ROI[8]. Setting the ROI as a constraint, an anchor is in homogeneously produced on the top feature map by the multitask RPN.

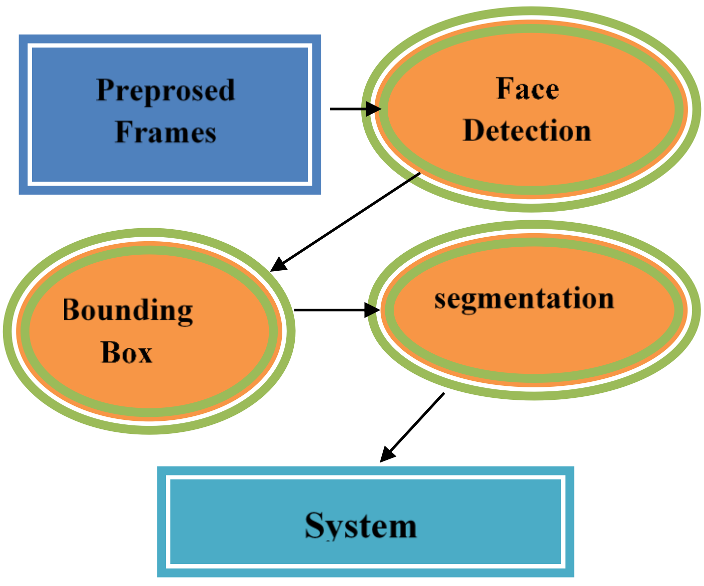

Fig. No.:3 Diagrammatic Representation of Face Expression and Detection

These are responsible for providing a predefined set of bounding boxes of different sizes and ratios that are going to be used for reference when first predicting object locations for the RPN.

4.2.4 Expression and Feature Extraction

- After acquiring the key frames of face expression, a deep feature representation of facial expressions is generated from a 2D-CNN.
- These deep learning models have layered architecture that learns features at different layers (hierarchical representations of layered features).
- This layered architecture allows extracting high-level, medium-level, and low-level features of face expression.
- Instead of acquiring features from just the last layer, features are extracted from convolution, pooling, and regularization layers.

4.2.5 Expression Classification

In Classification stage, Convolutional neural networks algorithm is used for classification of Face Expression images. It is a non-parametric method which is used for both classification and regression.

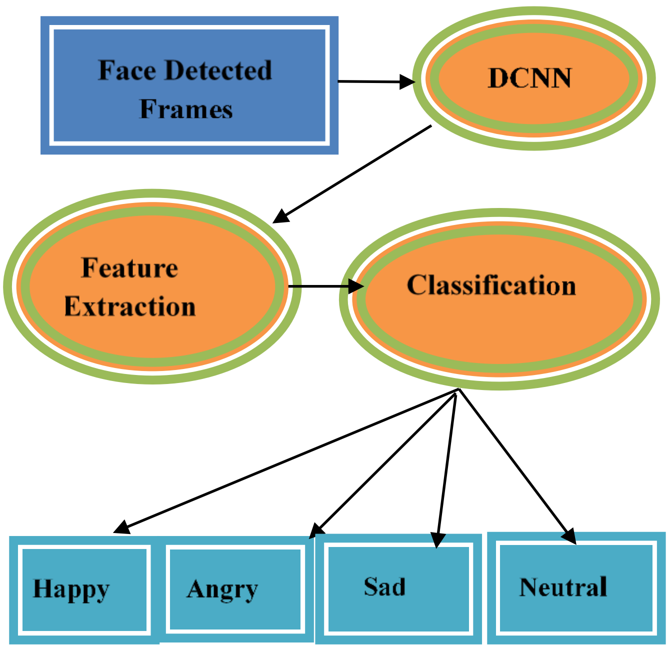

Fig.No.:4 Diagrammatic Representation of classification

Deep Convolution Neural Network Classifier: The Deep Convolution Neural Network (CNN) classifier is used mainly for image and video recognition. The CNN is able for automatically learning the respective feature for data itself. The CNN follows few steps like receiving different inputs, calculating the sum of their weights, forward output to activation function and respond with the desired output[9]. Based on CNN classification, the FER images important features like lines, edges, and object etc. complex features automatically able to identify with more accurately.

DCNN is used to emotion classification. First, DCNN classification model was constructed, and then the utterance-level features were used for DCNN training[10] and finally, the user produced output of DNN as the result of SER. DCNN has three layers, the first layer contains 512 cells with a dropout layer. And the second and third layer contains 256 cells, the output is one of the 6 classes emotion.

4.3 Facial Emotion Prediction-Testing Phase

4.3.1 Live Video Prediction

Cameras should be deployed in critical areas to capture relevant video. Computer and camera are interfaced and here webcam is used. For every participant, one video with six kinds of facial expressions is collected and processed[1]. A haar cascade classifier proposed by Viola and Jones is used to detect the face from video frame by frame. When a face is detected, the face image is saved into the database and labeled according to the facial expression the participant shows[2]. Because the database has a lot of similar images due to the successive frames, and uses the difference hash (dhash) algorithm to select representative images from the dataset.

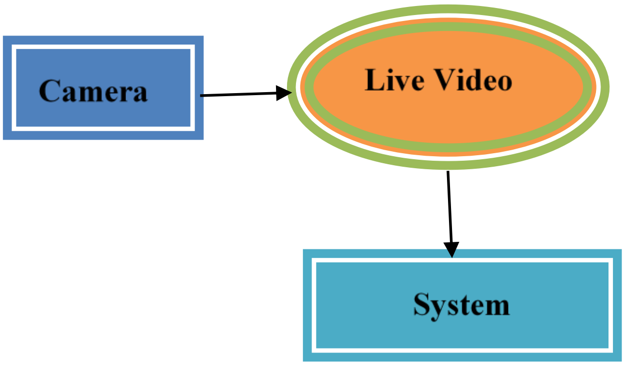

Fig.No.: 5 Diagrammatic Representation of Live video prediction

It’s important to note that the accuracy and performance of live video prediction for facial emotion prediction can vary depending on the quality of the video feed,lighting conditions,facial occlusions , and the robustness of the employed algorithms and models[3]. Additionally, it is crucial to have a well annotated training dataset that covers a wide range of emotions and includes diverse facial expressions to train accurate models for real-time prediction.

4.3.2 Predict Expression

In this module the matching process is done with trained classified result and test Live Camera Captured Classified file. Hamming Distance is used to calculate the difference according to the result the prediction accuracy will be displayed.

4.4 Recommendation

Facial Expression based recommend the health care systems where suggestions are based on an influence about a user’s emotion and based on a degree of domain expertise and knowledge. Rules are defined that set context for each recommendation

5. DATA FLOW DIAGRAM

A Data Flow Diagram (DFD) is a graphical representation of the "flow" of data through an information system, modeling its process aspects. Often they are a preliminary step used to create an overview of the system which can later be elaborated. DFD can also be used for the visualization of data processing.

A DFD shows what kind of information will be input to and output from the system, where the data will come from and go to, and where the data will be stored. It does not show information about the timing of processes, or information about whether processes will operate in sequence or in parallel.

A DFD consist of, The system captures real time video frames from a camera as input and then captured video frames are processed by a face detection algorithm to identify and locate faces in the frames.

Facial landmark detection detected faces are further analyzed to identify key facial landmarks, such as the position of the eyes, nose and mouth.The facial images containing the detected landmarks are preprocessed to enhance the quality and normalize the input data.This step may involve resizing, cropping, and applying image enhancement techniques. The preprocessed facial images are feed into a deep learning-based feature extraction model, such as a convolutional neural network (CNN), to extract relevant features that capture the facial expressions.

Emotion classification module inputted into a deep learning model, such as a fully connected neural network trained to specific emotion categories, such as happiness, sadness, anger, fear, surprise or disgust. Emotion output module is predicted emotions are generated based on the output of the emotion classification model. Refer below Fig.No.: 6

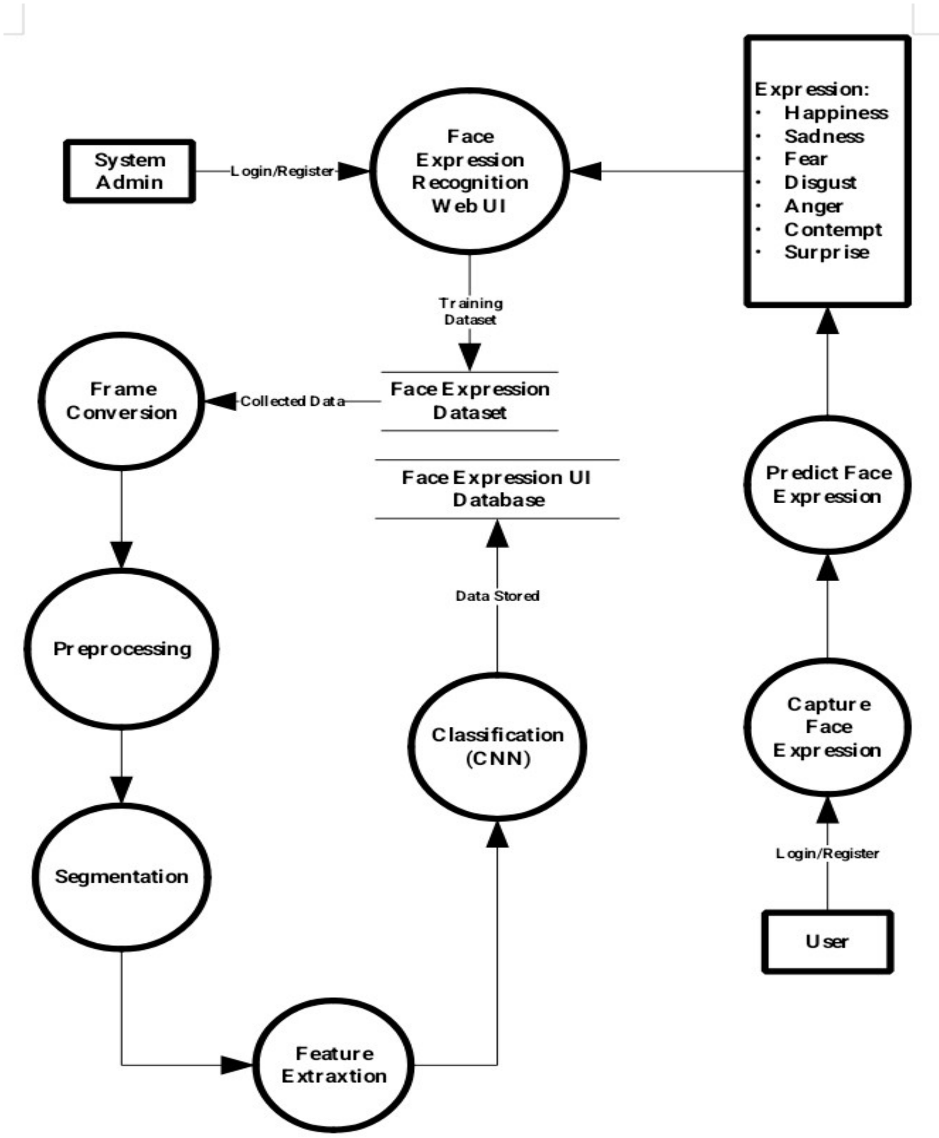

Fig. No.: 6 Data Flow Diagram

6. CONCLUSION

The author use a deep learning technique to process emotional big data and develop an emotion care system using facial expression recognition system.This project propose to use video to recognize emotional changes in continuous-time domain. In the field of artificial intelligence, emotion recognition has received a lot of attention.

Conventional emotion recognition algorithms distinguished emotion categories by detecting changes in facial expressions. Recently, various emotion recognition mechanisms based on convolutional neural network (CNN) which are trained in an end-to-end manner have been developed and showed reliable performance.

The Graphical Web User Interface allows users to do Realtime validation of the system. This project is considered seven discrete and unique emotion classes (angry, disgust, fear, happy, neutral, sad and surprise) for emotion classification. So, there is no overlapping among classes.

7. FUTURE ENHANCEMENT

- **Robustness to occlusions**

Facial expression recognition systems can be made more robust to occlusions, such as glasses or facial hair, by using data augmentation techniques.These techniques involve creating variations of the original image by applying transformations such as rotation, scaling, and cropping.

- **Multi-model input**

Facial expression recognition systems can be enhanced by incorporating other modalities, such as audio or body language, to improve the accuracy of facial expression recognition.For example, audio can provide additional cues to help identify the emotional state of a person.

- **Integration with Virtual Reality(VR):**

Integrate the facial expression analysis system with VR technologies to create immersive experiences that can track and respond to user's facial expressions within virtual environments.

- **Gesture and Body Language Recognition:**

Integrate additional modalities such as gesture and body language recognition to improve the overall understanding of users emotions and intensions.This could provide more comprenhensive insights into non-verbal commnunication

- **Mobile Application Development:**

Extend the system's accessibility by developing a mobile application. This would allow users to use the facial expression analysis system on their smartphones or tablets,enhancing portability and usability.

REFERENCES

[1] E. Ahmed, I. Yaqoob, I. A. T. Hashem, I. Khan, A. I. A. Ahmed, M. Imran, and A. V. Vasilakos, ``The role of big data analytics in Internet of Things," Comput. Netw., vol. 129, pp. 459 - 471, Dec. 2017.

[2] K. Lin, F. Xia, W. Wang, D. Tian, and J. Song, ``System design for big data application in emotion-aware healthcare," IEEE Access, vol. 4, pp. 6901-6909, 2016.

[3] Y. Jiang,W. Li, M. S. Hossain, M. Chen, A. Alelaiwi, and M. Al-Hammadi, ``A snapshot research and implementation of multimodal information fusion for data-driven emotion recognition," Inf. Fusion, vol. 53, pp. 209 - 221, Jan. 2020.

[4] A. Chatterjee, U. Gupta, M. K. Chinnakotla, R. Srikanth, M. Galley, and P. Agrawal, ``Understanding emotions in text using deep learning and big data," Comput. Hum. Behav., vol. 93, pp. 309317, Apr. 2019.

[5] Y. Ma, Y. Hao, M. Chen, J. Chen, P. Lu, and A. Ko²ir, ``Audio-visual emotion fusion (AVEF): A deep efficient weighted approach," Inf. Fusion, vol. 46, pp. 184 - 192, Mar. 2019.

[6] M. S. Hossain and G. Muhammad, ``Emotion recognition using deep learning approach from audio-visual emotional big data," Inf. Fusion, vol. 49, pp. 69 - 78, Sep. 2019.

[7] Y. Miao, H. Dong, J. M. A. Jaam, and A. E. Saddik, ``A deep learning system for recognizing facial expression in real-time," ACM Trans. Multimedia Comput., Commun., Appl., vol. 15, no. 2, pp. 1 - 20, Jun. 2019, doi: 10.1145/3311747.

[8] G. Muhammad, M. S. Hossain, and N. Kumar, ``EEG-based pathology detection for home health monitoring," IEEE J. Sel. Areas Commun., vol. 39, no. 2, pp. 603 - 610, Feb. 2021.

[9] A. E. Saddik, ``Digital twins: The convergence of multimedia technologies," IEEE Multimedia, vol. 25, no. 2, pp. 8792, Apr./Jun. 2018.

[10] M. Chuah and M. Diblasio, ``Smartphone based autism social alert system," in Proc. 8th Int. Conf. Mobile Ad-hoc Sensor Netw. (MSN), Dec. 2012, pp. 6 - 13.