IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

FACIAL EXPRESSION RECOGNITION BASED ON CONVOLUTIONAL NEURAL NETWORK

[1]SHAH DHRUVIL AMIT, [2] SHETTY SATHVIK RAVINDRA, [3] SHETTY DISHA RAVINDRA,

[4] SHETTY ANKIT SURESH, [5]VIDYA

[1]Student, [2]Student, [3]Student, [4]Student, [5]Assistant Professor

[1,2,3,4,5] Department of Computer Science and Engineering

[1,2,3,4,5] Alva's Institute of Engineering and Technology, Mijar, India

Abstract: Human emotions are natural expressions that people make without any conscious effort that is accompanied by the reflexing of facial muscles. Common emotions that a human face expresses in response to various situations include happiness, sadness, surprise, anger, and stable normal. The proposed software detects and recognizes faces as well as recognizes a lot more about that human, which could be used to get feedback from customers or to determine if a person requires assistance. The project's main goal is to create a product that is both affordable and efficient. The AI and digital image processing techniques are used to create the system in Python. Drowsiness is important in situations where there is a need to avoid accidents or mishaps, such as driving or security vigilance. As for the system's ability to recognize the identity card, it is a simple feature in which the installed camera is trained to first focus on the card and recognize its shape and color.

Index Terms **– CNN Model, Emotion Detection, Image Processing**

I. INTRODUCTION

The field of Artificial Intelligence and Digital Image Processing is growing in our country slowly and steadily with the majority related to face recognition. Several areas of the industry have started using the various techniques and applications of AI and DIP. The project can be implemented for marketing purposes also, as it lets us know the feedback of any product. It provides accurate results and is easy to implement and understood in the most common systems. Also, these features can be installed cost-effectively and efficiently in schools or colleges, or any other area where surveillance is required but lack of finances is a major factor. So, using our proposed project, surveillance could be provided, which results in help in maintaining a regular health check and understanding the emotions of a person in the workplace. It can also be used as feedback by workers after making some changes in the workplace. Artificial Intelligence & Digital image Processing technology is used to make the system which contains face recognition, emotion recognition; drowsiness detection, and id-card detection. In face recognition, the CNN algorithm is used. The given proposed work has shown us that the performance of the face recognition technology can be improved much better by mixing Gabor wavelet and LBP for features extraction for classification. We can understand a person's emotions if we can analyse them at different stages. For this purpose, we aim to develop a Convolutional-Neural Network (CNN) based on the Facial Expression Recognition System (FER). The algorithm used for drowsiness detection detects the blinking of the eye through the camera installed in the system that predicts the output. The practice of detecting human emotions from facial expressions is known as facial emotion recognition. The human brain perceives emotions instinctively, and software that can recognize emotions has recently been developed. This technology is constantly improving, and it will eventually be able to detect emotions as precisely as our brains. By learning what each facial expression signifies and applying that knowledge to fresh information, AI can recognize emotions. Emotional artificial intelligence, often known as emotional artificial intelligence, is a type of artificial intelligence that can read, imitate, interpret, and respond to human facial expressions and emotions is known as artificial intelligence.

Over the last two decades, facial expression recognition (FER) has become a popular study topic. Humans utilize facial expressions to quickly, naturally and effectively communicate their intentions and sentiments. Many major applications for the FER system include driver safety, health care, video conferencing, virtual reality, and cognitive research, among others.

Generally, facial expressions can be classified into neutral, anger, disgust, fear, surprise, sad, and happy. Recent research shows that the ability of young people to read the feeling and emotions of other people is getting reduced due to the extensive use of digital devices. As a result, it's vital to create a FER system that can reliably recognize facial expressions in real-time. Preprocessing, feature extraction, feature selection, and facial expression categorization are the four processes of an automatic FER system. In the preprocessing step, the face region is first detected and then extracted from the input image because it is the area that contains expression-related information.

II. LITERATURE SURVEY

[1] Data mining, which is also known as data or knowledge discovery, is the process of analysing data through different perspectives and summarizing it into useful information for further use. Image processing is related to computer vision, which is high-level image processing out of which software intends to decipher the physical contents of an image. Out of multiple ways, one of the ways to do this is by comparing selected facial features from the image and a facial database. The task of recognizing emotion from images has become one of the active research themes in image processing and in applications based on human-computer interaction, which is taken care of by many researchers. This research conducted an experimental study on recognizing facial emotions. The flow of our emotion recognition system includes the basic process in the FER system. This includes image procurement, pre-processing of an image, face detection, feature extraction, classification of emotions, and then when the emotions are classified, the system assigns the person with particular music depending on the emotions. The system here focuses on live images taken from the webcam. The main aim here is to develop an automatic facial emotion recognition system for stressed individuals, hence assigning them to music therapy to relieve stress. Happiness, sadness, surprise, fear, disgust, and anger are all commonly recognized emotions that were considered in the trials.

[2] Human-Computer Interaction requires emotion recognition from voice signals, which is a difficult task (HCI). Many strategies have been used in the field of speech emotion recognition (SER), extracting emotions from signals, including much well-established speech analysis and classification techniques that are already well-established. Deep Learning approaches have lately been presented as a substitute to standard techniques in SER. The author proposed an overview of Deep Learning techniques and discusses some recent literature where these methods are utilized for speech-based emotion recognition. The review done here goes over databases used, emotions extracted, contributions made toward speech emotion recognition, and limitations related to it.

[3] There are various applications using body sensor networks (BSNs) that constitute a new trend in car safety. Furthermore, detecting human behavioural states that may affect driving poses a significant difficulty when using heterogeneous body sensors and vehicle ad hoc networks (VANETs). This paper proposes the detection of various human emotions, from which emotions such as tiredness (drowsiness) and stress (tension) are extracted, which could be a major reason for traffic or accidents. They present an exploratory study demonstrating the feasibility of detecting one emotional state in a real-time environment using a BSN. The results obtained here gave a basis to propose a middleware architecture that is capable of detecting emotions, which can be communicated through the onboard unit of a vehicle to various entities such as city emergency services, VANETs, and roadside units, with the purpose of improving the driver's experience and also guaranteeing better security measures in order to provide road safety.

[4] A real-time algorithm to detect eye blinks in a video stream from a standard camera or a webcam is proposed in this paper. Recent landmark detectors trained on in-the-wild datasets exhibit excellent robustness against a camera's head position, as well as fluctuating illumination and face expressions across time. They show that the landmarks are detected precisely enough to reliably estimate the level of the eye-opening in the video sequence. The proposed algorithm, therefore, estimates the landmark positions throughout the video, extracts a single scalar quantity – eye aspect ratio (EAR) – characterizing the eye-opening in each frame. Finally, a Support Vector Machine (SVM) classifier detects eye blinks as a pattern of EAR values in a short temporal window. The simple algorithm proposed here outperforms the state-of-the-art results on two standard datasets.

[5] One of the most difficult aspects of face recognition is dealing with differences in orientation or position, lighting variations, facial expressions, occlusions, and aging. They present a method for face identification in an uncontrolled environment in this work that combines Gabor wavelets with Local Binary Patterns during the feature extraction phase. Then, we apply the dimension reduction technique to reduce the pattern vectors in the extracted features. Finally, we combine both K Nearest Neighbour (KNN) and Sparse Representation Classifier for the face recognition phase. On the basis of the LFW database, we evaluated our method and conducted comparative experimental research involving many experiments. With a recognition rate of 94 percent, the best result is obvious.

III. REQUIREMENTS ANALYSIS

In the autonomous vehicle sector, computer-assisted learning is a rapidly increasing and dynamic area of research. The recent researchers in machine learning and artificial intelligence promise the improved accuracy of perception of emotion and drowsiness detection. Here, computers are enabled to think by developing intelligence by learning. There are many types of Machine Learning Techniques that are used to classify data sets.

3.1 Functional Requirements

The proposed application should be able to identify input face emotion and detect emotion type. The function of a system and its components are defined by functional requirements. A function is described as a set of inputs, the behavior, and its outputs.

Functionality

The application is developed in such a way that any future enhancement can be easily implementable. The maintenance required for this particular project is very minimal. The software used for the development is open source and can be installed easily. The application is developed in such a way that it is easy to use and install for any person of interest.

Reliability

It is maturity, fault tolerance, and recoverability. The system is reliable for any number of user input and training datasets. The emotion dataset is taken for emotion recognition and the Haar-Cascade classifier is used for eye drowsiness detection.

Usability

It is easy to understand, learn and operate the software system. The user can show the face and learn the appropriate emotion.

Safety

Safety-critical issues associated with its integrity level. The computer system being used is protected by a password.

Security

Some accessible ports are not blocked by the Windows firewall. The web camera port should be enabled automatically, otherwise, the user must enable it every time.

Communications

The application is developed in such a way that communication can be handled through the camera. Similarly, emotion detected is printed on the screen through live web camera detection.

3.2 Non-Functional Requirements

Non-functional requirements determine the resources required, time interval, transaction rates, throughput, and everything that deals with the performance of the system.

Maintainability

It is easy to maintain the system as it does not require any special maintenance after download. Updates are required only if notified to the user about any. Easy maintenance is one of the features that make this proposal most usable.

Portability

The software must easily be transferred to another environment, including installability. It can be carried about as simple as a standard computer. The user can access the computer from the place where the system was installed.

Performance

Less time for detection of signs once the input has arrived. Similarly, the training time is also less as we are given a limited epoch on training.

Accuracy

The accuracy generated by our work is outperformed by any other existing models. We can recognize emotions and eye drowsiness accurately through our proposed system.

3.3 Feasibility Study

The project's feasibility is examined in this phase, and a business proposal with a high-level project plan and cost estimates is provided. During system analysis, the feasibility study of the proposed system is carried out, which ensures that the system to be proposed is not a burden to the company. A basic understanding of the system's primary requirements is required for feasibility analysis.

Technical Feasibility

This study was conducted to determine the system's technical feasibility or technical requirements. Any system that is created should not place a large burden on the available technical resources. As a result, there will be a lot of demand for the available technical resources. As a result, the client will be subjected to severe demands. Because very minor or no changes are necessary to implement this system, the designed system must have a low requirement.

Economic Feasibility

This study was carried out to check the economic impact that the system will have on the organization. The amount of money the corporation has to invest in the system's research and development is limited. It is necessary to justify the spending. As a result, the final product came in under budget, thanks to the fact that the majority of the technologies used were freely available. The only things that needed to be purchased were the customized ones.

Social Feasibility

The purpose of the study is to establish the system's level of acceptance among users. This covers the process of teaching the user how to effectively use the technology. Instead of being fearful of the system, the user should accept it as a necessity. The methods used to educate and familiarise the user with the system are totally responsible for the level of acceptance by the users.

3.4 Software Specification

The project is primarily concerned with the detection of sign languages. We implemented it with the Python 2.6 version. The libraries required are to be installed prior to executing the project without any hurdles. We installed CV2 for OpenCV, Kera's, TensorFlow, NumPy, etc.

Hardware Requirements

Processor: Any Processor above 500 MHz

Ram: 4 GB

Hard Disk: 250 GB

Input devices: Standard Keyboard and Mouse, Web Camera

Output device: High Resolution Monitor.

Software Specification

Operating System: Windows 7 or higher

Programming: Python 3.6 and related libraries

IV. METHODOLOGY

Hardware Specifications:

1. Processor: Any processor with a clock speed greater than 500 MHz.
2. RAM: 4 GB
3. Hard Drive: 250 GB
4. Input Devices: Standard Keyboard and Mouse, as well as a Web Camera
5. High Resolution Monitor as an output device

Software Requirements:

1. Windows 7 or higher as the operating system

2. Python 3.6 and related libraries for programming

The following modules are included:

- Collection of datasets
- Pre-processing of images
- Convolutional 2D neural network training
- The detection of eyes
- Recognition

The data collected includes grayscale images of faces in the size of 48x48 pixels. The faces are automatically considered so that the face is nearly in the center and occupies roughly the same amount of space in each image. Faces are classified into one of seven classes based on the emotion displayed in their expressions (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The file fer2013.csv has two columns: "emotion" and "pixels." For the emotion depicted in the image, the column for "emotion" has a numeric code ranging from 0 to 6, inclusive. For each image, the "pixels" column contains a string with quotes. This string contains space-separated pixel values that are mostly distributed in rows test. The columns for "pixels" are contained in a csv file, and the task is to predict the column for emotion.
There are 28,709 examples in the training set. There are 3,589 examples in the set of public tests used for the leaderboard. The final test set helped to determine the winner of the competition, which included another 3,589 examples'-PROCESSING OF IMAGES
Image pre-processing consists of several steps, such as color conversion.
To convert the input image to grey scale, we used the BGR2GRAY function, one of many color conversion functions (to convert input image from one color space to another).

CONVOLUTIONAL 2D NEURAL NETWORK TRAINING

We trained and tested our model using a convolutional 2F neural network available in Kera's. Conv2D's overall architecture is depicted below.

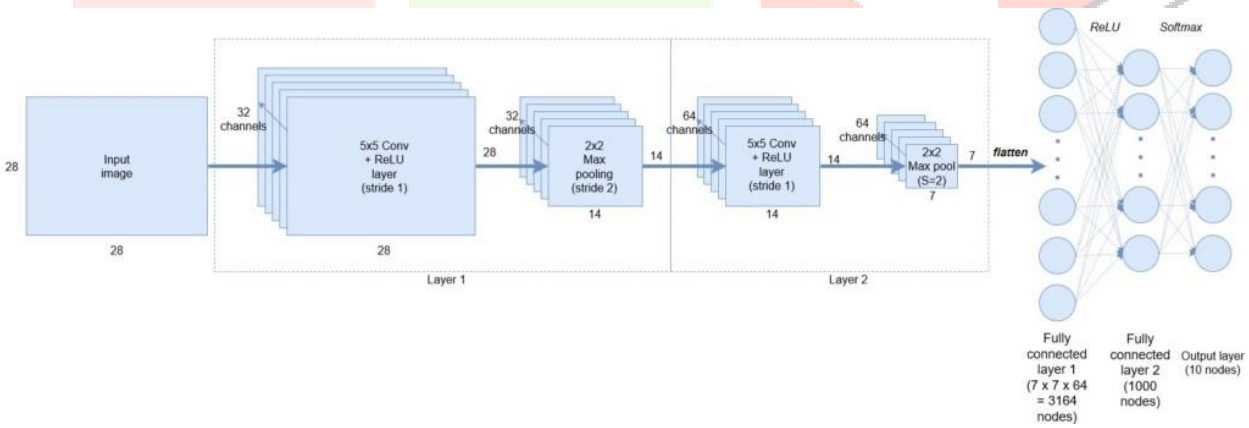

Fig 1: CNN Model

Model of Sequence

Kera's models are classified into two types: sequential and functional. The Sequential model is most likely for most deep learning networks. It allows you to arrange the network's sequential layers alongside its recurrent layers in the order of input to output.
The model type is declared as Sequential in the first line (). Adding 2D Convolutional layer.

To process the 2D input images, add a 2D convolutional layer. The first argument to the Conv2D () layer function is the number of output channels, which in this case is 32. The second input is the kernel size, which we have set to a 55-frame moving window, followed by the x and y strides (1, 1). The activation function is then a rectified linear unit, and we must finally supply the model with the size of the input to the layer. The first layer must declare the input shape – Kera's assists in determining the size of the tensors that flow through the model from there. Including a 2D max pooling layer, Add a pooling layer in 2D max. In this case, we specify the pooling size in the x and y directions, as well as the strides including a new convolutional + max pooling layer.

Then, with 64 output channels, we add another convolutional + max pooling layer. In Keras, the default argument for strides in the Conv2D() function is (1, 1). Keres's default strides argument is to make it equal to the pool size. This layer's input tensor is (batch size, 28, 28, 32) – 28x28 is the image size, and 32 is the number of output channels from the previous layer.
Adding a dense layer and flattening
The output from these must be flattened before it can enter our fully connected layers. The next two lines declare our fully connected layers – we specify the size using Kera's' Dense () layer – in accordance with our architecture, we specify 1000 nodes, each activated by a ReLU function. The size of the number of classes is determined by our soft-max classification, or output layer.Training neural network.

We must specify the loss function or tell the framework what type of optimizer to use in the training model (i.e. gradient descent, Adam optimizer etc.).

For categorical class classification, use the Lass function of standard cross entropy (keras.losses.categorical cross entropy). We employ the Adam optimizer (keras. optimizers. Adam). Finally, we must specify a metric that will be computed when evaluate () is applied to the model. To begin, enter all of the training data – in this case, x train and y train. The batch size is the next argument. In this case, the batch size is 32. Following that, we pass the number of training epochs (2 in this case). There is a verbose flag, which is set to 1 in this case, indicating if you want detailed information printed on the console.

DETECTION OF EYE DROWSINESS

The haarcascade classifier, "haarcascade eye.xml," is used for eye detection. An eyeball would most likely not be discovered by eye detection. Eye detection entails looking for eye features such as surrounding skin, lids, lashes, and brows. The eye size is detected for openness or closure, and an alert is generated on the output.

RECOGNITION

Finally, the image or web camera input is passed. Predictions are made for emotion and eye drowsiness. The predict.py module defines emotion recognition and drowsiness.

V. SYSTEM TESTING

After completing software development, the next complicated and time-consuming step is software testing. Only the development team knows how far the user requirements have been met during this process, and so on. This phase ensures software quality and provides the final review of specification, design, and coding. The increasing convenience of software as a system, as well as the cost associated with software failures, are driving forces for thorough testing.

5.1 Objectives of Testing: These are the rules that account for testing objectives:

- Testing is the process of running a programme in order to find errors.
- A good test case is one that has a high chance of detecting an undiscovered error.
- The project's testing procedures are carried out in the order listed below.
- System testing is performed to verify the server's name of the machines that are linked between the customer and the executive.
- The product information provided by the company to the executive is validated using the centralised data store.
- System testing is also performed to ensure that the executive is available to connect to the server.
- The server's name authentication is checked, as is the customer's availability.
- Proper communication chat line viability is tested, and the chat system is made to work properly.
- Mail functions are validated against user concurrency and customer mail date.

5.2 TEST PERFORMED

Some of the testing methods used in this successful project are listed below:

S.NO	Test Case ID	Test Description	Test Procedure	Test Input	Expected Result	Actual Result
1	T101	To check training	Start training dataset	Execute train.py	Training should start and created output files	Trained weight file in h5 format is created
2	T102	To check image prediction	Image input	Execute predict.py	Prediction should start and output will come for image	Successfully completed
3	T104	To check recognition	Start recognition by input video or video	Execute recognize.py	Output of corresponding emotion value	Human Emotion + Drowsiness

Table 1: Test performed using the system

VI. CONCLUSION

The proposed work uses image and web camera data to detect driver drowsiness. The proposed work detects both emotion and drowsiness. This aids in driver monitoring while driving as well as emotional monitoring. Continuous research is being conducted in order to improve on the existing ones. The study field of Artificial Intelligence and Digital Image Processing is expanding at an exponential rate, so the future prospects are extremely promising. Many methods have already been proposed, but improved versions are on the way. This enables the machine to determine what actions are required in specific situations. It is as simple as it appears, but the efforts required are extremely complex. Training machines to think like humans can be a difficult task, but the research and other related work that has been done so far has been fantastic, and it will only get better in the future as computer science advances.

REFERENCES

[1] Deshmukh, Renuka & Paygude, Shilpa & Jagtap, Vandana. (2017). Facial Emotion Recognition System through Machine Learning approach. 10.1109/ICCONS.2017.8250725.

[2] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar and T. Alhussain, "Speech Emotion Recognition Using Deep Learning Techniques: A Review," in IEEE Access, vol. 7, pp. 117327-117345, 2019.

[3] Rebolledo-Mendez, Genaro & Reyes, Angelica & Paszkowicz, Sebastian & Domingo, Mari & Skrypchuk, Lee. (2014). Developing a Body Sensor Network to Detect Emotions During Driving. Intelligent Transportation Systems, IEEE Transactions on. 15. 1850-1854.

[4] Tereza Soukupov´a and Jan Cech., 21st Computer Vision Winter Workshop Luka Cehovin, Rok Mandeljc, Vitomir Struc (eds.) Rimske Toplice, Slovenia, February 3–5, 2016

[5] B. Ameur, S. Masmoudi, A. G. Derbel and A. Ben Hamida, "Fusing Gabor and LBP feature sets for KNN and SRC-based face recognition," 2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Monastir, 2016, pp. 453-458.