Control Abnormal Situations with Facial Expressions Using Deep Learning Techniques

[1] A. Sathi Babu, [2]P.N.B. Swamy, [3]CH. Sai Sindhu, [4]G. Preethi, [5]B. Prudhvik, [6]G. Ashok.
[2, 3, 4.5] *B.Tech In Dept. of ECE, NRI Institute of Technology Vijayawada Andhra Pradesh 521212*
[1,2] *Assistant Professor, Dept. of ECE, NRI Institute of Technology Vijayawada Andhra Pradesh 521212*

Abstract: In this model we give the overview of the work done in the past related to Emotion Recognition using Facial expressions along with our approach towards solving the problem. Based on the expression detected by the system it displays the output about the problems it may lead to like for example, if the system detects that the person is angry it displays a warning that your blood pressure may increase and it may deteriorate your health.

The system operates under two different conditions one when the system is having internet access and other when it doesn't have internet access. When the system is having internet access, based on different expressions recognised the system displays a precaution and redirects to different links based on different expressions like For Example, if the system detects that the person is sad then it displays a warning which says your sadness may lead to depression and suggests you to listen songs and redirects you to Spotify page to play some soothing music. And when the system doesn't have internet access, it detects the expression and displays a warning based on your expression.

Keywords: Blood pressure, Depression, Emotion Recognition, Precautions.

INTRODUCTION

Emotion recognition is a process of identifying the human emotions most likely from facial expressions as well as from speech. The application of emotion recognition system is that it promotes emotion translation between cultures that can be used in multi-cultural communication systems. After extensive research, it is now generally accepted that humans share seven facial expressions that reflect the experiencing of fundamental emotions. These fundamental emotions are anger, neutral, disgust, fear, happiness, sadness, and surprise. Human beings have the capability to recognize emotions easily, but it is difficult for the computers to do the same. If computers could recognize these emotional inputs, they could give specific and appropriate help to users in ways that are more in tune with the user's needs and preferences. It can be used to find application where efficiency and automation can be useful, including in entertainment, social media, content analysis, criminal justice, and healthcare. For example, content providers can determine the reactions of a consumer and adjust their preferences accordingly. Facial expressions help computers in detecting emotions. This paper deals with helping computers to recognize human emotions in real-time.

A Facial expression is the visible manifestation of the affective state, cognitive activity, intention, personality and psychopathology of a person and plays a communicative role in interpersonal relations. Human facial expressions can be easily classified into 7 basic emotions: happy, sad, surprise, fear, anger, disgust, and neutral. Our facial emotions are expressed by activation of specific sets of facial muscles. These sometimes subtle, yet complex, signals in an expression often contain an abundant amount of information about our state of mind. Automatic recognition of facial expressions can be an important component of natural human machine interfaces; it may also be used in behavioural science and in clinical practice. It have been studied for a long period of time and obtaining the progress recent decades. Though much progress has been made, recognizing facial expression with a high accuracy remains to be difficult due to the complexity and varieties of facial expressions.

LITERATURE SURVEY

In a paper [1], a hybrid approach in which multi modal information for facial emotion recognition is used. In the experiment conducted by authors, they chose two

different speakers using two different languages. The evaluation is carried out with three different media clips, (1) audio information of the emotions only, (2) video information of the emotions only, (3) both audio and video information (original video clip). Video and audio dominance of each type of emotion is recorded and compared. The results of audio and facial recognition are provided as input to the weighing matrix. Inside the weighing matrix computations are made and the expression whose computed value is maximum is the result. According to a paper [2], the problem that was solved is about Emotion recognition using facial expression. Microsoft Kinect was used for 3D modelling of the face. Microsoft Kinect has 2 cameras. One works with visible light and the other one works with infrared light. It gives three-dimensional co-ordinates of specific face muscles. Facial Action Coding System (FACS) was used to return special coefficients called Action Units (AU). There are 6 Action Units. These Action Units (AU) represent different region of face. Six men of the age group 26-50 participated and tried to mimic the emotions specified to them. Each person had 2 sessions and each session had 3 trials. 3-NN had an accuracy of 96%.MLP had an accuracy of 90%. According to the paper [3], CERT can detect 19 different facial actions, 6 different prototypical emotions and 3D head orientation using Facial Action Unit Coding System (FACS) and three emotion modules. It follows 6 stages: (1) Face Detection using Gentle Boost as boosting algorithm, (2) Facial Feature Detection – Specific location estimates are estimated by combining log likelihood ratio and feature specific prior at that location, and these location estimates are refined using Linear regressor, (3) Face Registration – affine wrap is made and L2 Norm is minimized between wrapped facial feature position and canonical position from GENKI dataset, (4) Feature Extraction – feature vector is obtained using Gabor filter on face patch from previous patch, (5) Action Unit Recognition – feature vector is fed to Support Vector machine to obtain Action Unit Intensities, (6) Expression Intensity and Dynamics – Empirically CERT outputs significantly correlates with facial actions. In a paper [4], Psychological theories state that all human emotions can be classified into six basic emotions: sadness, happiness, fear, anger, neutral and surprise. Three systems were built- one with audio, another with face recognition and one more with both. The performances of all the systems were compared. Features used for speech-global prosodic features, for facedata from 102 markers on face. Both feature level and decision level integration were implemented. The result proved that performance of both the systems was similar. However, recognition rate for specific emotions presented significant errors. The type of integration to be used is dependent on the nature of the application.

METHODOLOGY

The facial expression recognition system is implemented using convolutional neural network. The block diagram of the system is shown in following figures, during training, the system received a training data comprising grayscale images of faces with their respective expression label and learns a set of weights for the network. The training step took as input an image with a face. Thereafter, an intensity normalization is applied to the image. The normalized images are used to train the Convolutional Network. To ensure that the training performance is not affected by the order of presentation of the examples, validation dataset is used to choose the final best set of weights out of a set of trainings performed with samples presented in different orders. The output of the training step is a set of weights that achieve the best result with the training data. During test, the system received a grayscale image of a face from test dataset, and output the predicted expression by using the final network weights learned during training. Its output is a single number that represents one of the seven basic expressions.

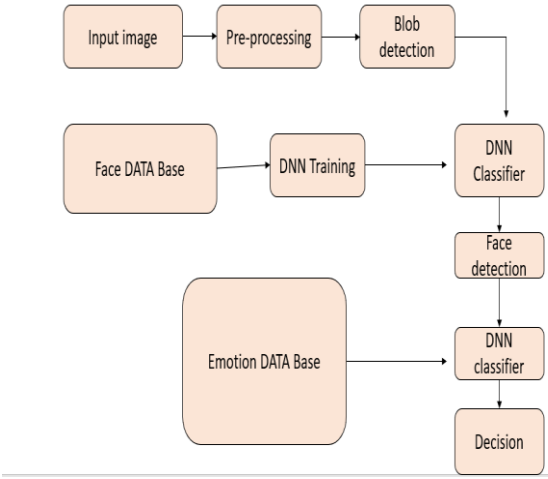
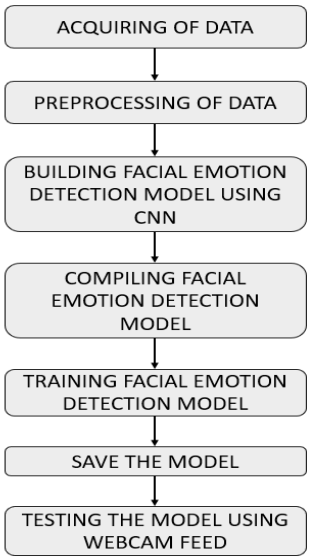


Fig: Block diagram for proposed system.

FLOW CHART



STAGES OF COMPUTATION

Dataset

The dataset from a Kaggle Facial Expression is used for the training and testing. It is a csv file consisting of image in pixels (where each pixel varies from 0 - 255) along with a label indicating the emotion of the person in image. It comprises pre-cropped, 48-by48-pixel grayscale images of faces each labelled with one of the 7 emotion classes: anger, disgust, fear, happiness, sadness, surprise, and neutral. Dataset has training set of 35,887 facial images with facial expression labels. The dataset has class imbalance issue, since some classes have large number of examples while some has few. The dataset is balanced using oversampling, by increasing numbers in minority classes. The balanced dataset contains 40263 images, from which 29263 images are used for training, 6000 images are used for testing, and 5000 images are used for validation.

Pre-processing

Firstly, we collected the dataset from the public domain containing images of facial expression of different person. These images collected from the public domain are raw images. These raw images are to be processed before they are ready to use. Each image is different from other image based on size. More uniformity in data means more accuracy of the model. Therefore, we pre-process the data by rescaling or normalizing the data.

TRAINING DATA AND TEST DATA ARE TWO IMPORTANT CONCEPTS IN MACHINE LEARNING

1.Training Data: The observations in the training set form the experience that the algorithm uses to learn. In supervised learning problems, each observation consists of an observed output variable and one or more observed input variables. 2) Test Data: The test set is a set of observations used to evaluate the performance of the model using some performance metric. It is important that no observations from the training set are included in the test set. If the test set does contain examples from the training set, it will be difficult to assess whether the algorithm has learned to generalize from the training set or has simply memorized it A program that generalizes well will be able to effectively perform a task with new data. In contrast, a program that memorizes the training data by learning an overly complex model could predict the values of the response variable for the training set accurately, but will fail to predict the value of the response variable for new examples. Memorizing the training set is called over-fitting. A program that memorizes its observations may not perform its task well, as it could memorize relations and structures that are noise or coincidence. Balancing memorization and generalization, or over-fitting and under-fitting, is a problem common to many machine learning algorithms. Regularization may be applied to many models to reduce over-fitting.

Classification

Previously built Convolution Neural Network (CNN) model is used. The trained data is used to predict emotion. A list with probabilities of all 7 emotions is obtained as an output. The required output is the maximum of these values and the corresponding emotion is predicted as the final output.

ARCHITECTURE OF CNN

Convolutional Neural Networks (ConvNets or CNNs) are a category of Neural Networks that have proven very effective in areas such as image recognition and classification. ConvNets have been successful in identifying faces, objects and traffic signs apart from powering vision in robots and self-driving cars. Convolutional neural network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied

to analysing visual imagery. CNNs are regularized versions of multilayer perceptrons. Multilayer perceptron usually mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. Convolutional networks were inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. Individual cortical neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visual field. A typical architecture of a convolutional neural network contains an input layer, some convolutional layers, some fully-connected layers, and an output layer. CNN is designed with some modification on LeNet Architecture. It has 6 layers without considering input and output. The architecture of the Convolution Neural Network used in the project is shown in the following figure.

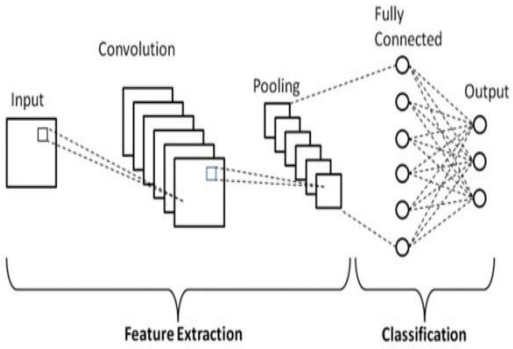


Fig: ARCHITECTURE OF CNN

RESULTS

OUTPUTS:

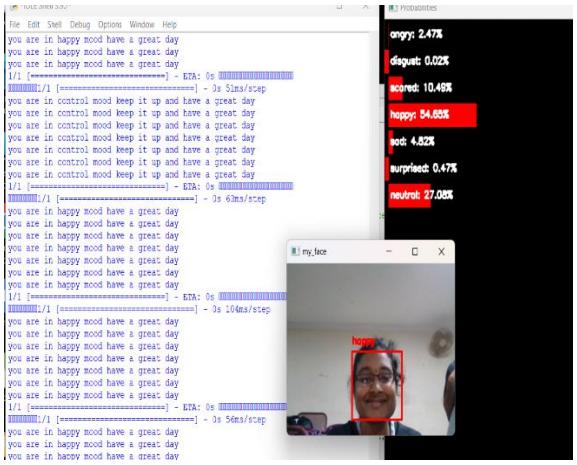


Fig:

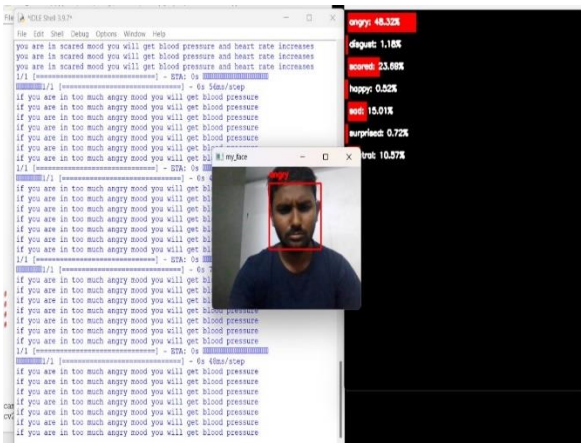


Fig:

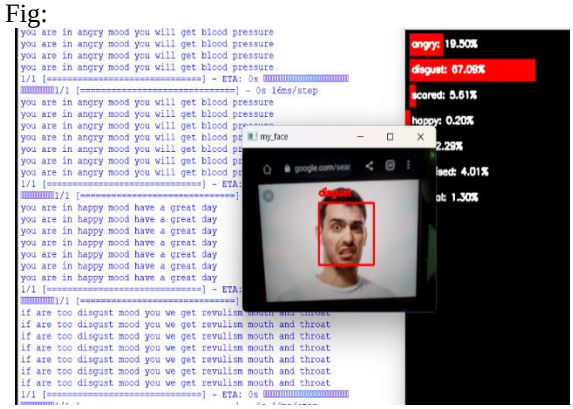


Fig:

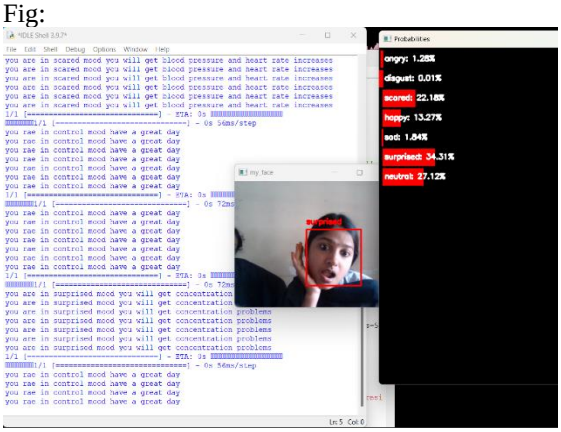


Fig:

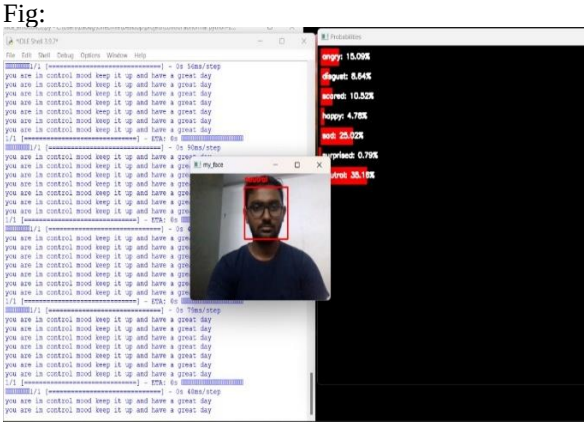


Fig:

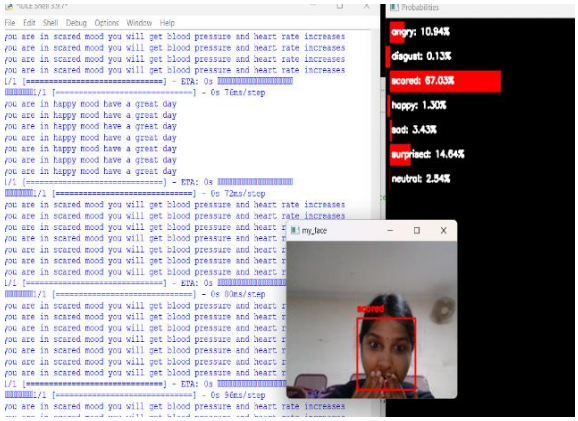


Fig:

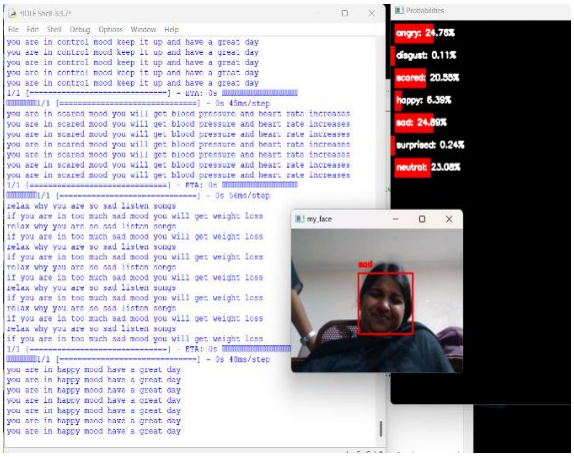


Fig:

For future work, our system's precision and accuracy can be improved by collecting more data from different sources & training our model with it. Additionally, in the future we can use different techniques to extract more features from the images or dataset. So that, our model can very easily detect the proper emotions for all the faces. Nowadays, emotion recognition is used for various purposes that some people do not even notice on a daily basis. Here are few areas where emotion recognition is used: A. Security Measures Emotion recognition is can used by schools and other institutions since it can help prevent violence and improves the overall security of a place. B. HR Assistance There are companies that use AI with emotion recognition system as HR assistants. The system is helpful in determining whether the candidate is honest and truly interested in the position by evaluating intonations, facial expressions, keywords, and creating a report for the human recruiters for final assessment. C. Customer Service There are companies that use AI with emotion recognition system as HR assistants. The system is helpful in determining whether the candidate is honest and truly interested in the position by evaluating intonations, facial expressions, keywords, and creating a report for the human recruiters for final assessment. D. Differently Abled Children There is a project using a system in Google smart glasses that aims to help autistic children interpret the feelings of people around them. When a child interacts with other people, clues about the other person's emotions are provided using graphics and sound. E. Audience Engagement Companies are also using emotion recognition to determine their business outcomes in terms of the audience's emotional responses. Apple also released a new feature in their iPhones where an emoji is designed to mimic a person's facial expressions, called Animoji. F. Video Game Testing Video games are tested to gain feedback from the user to determine if the companies have succeeded in their goals. Using emotion recognition during these testing phases, the emotions a user is experiencing in real-time can be understood, and their feedback can be incorporated in making the final product. G. Healthcare The healthcare industry sure is taking advantage of facial emotion recognition nowadays. They use it to know if a patient needs medicine or for physicians to know whom to prioritize in seeing first.

CONCLUSION AND FUTURE SCOPE

This project recognises the facial expressions and emotions by using deep learning techniques. In this we proposed real time video surveillance, what human face expresses it in front of camera and they were recognising the emotions after recognising the system also displays the disclaimer based on different expressions exhibited by a person. In this paper, an image processing and classification method has been implemented in which images of the faces are used to train a classifier predictor that predicts the seven basic human emotions for the given test images. The procedure to predict the emotions of a person by processing the image which was taken by web cam through various stages, such as pre-processing, face detection, and classifier using CNN is showcased.

In future we have increase the accuracy rate based on abnormal detection using hardware with music.

REFERENCE

[1] H. L. Wagner, R. Buck, M. Winterbotham, "Communication of specific person behaviors: Gender differences in sending accuracy and communication measures," J. Nonverbal. Behav., vol.17, pp.29-52, 1993.

[2] J. G. Allen, D. M. Haccoun, "Sex differences in person behaviorality: A multidimensional approach," Hum. Relat., vol. 29, pp. 711- 722, 1976.

[3] C. L. Barr, R. E. Kleck, Self-other perception of the intensity of facial expressions of person behavior: Do we know what we show," J. Pers. Soc. Psychol.vol. 68, pp.608-618, 1995.

[4] A. M. Kring, A. H. Gordon, "Sex differences in person behavior: Expression, experience, and physiology," J. Pers. Soc. Psychol. vol. 74, pp. 686-703, 1998.

[5] L. R. Brody, Gender, person behavioral expression, and parent-child boundaries. In R. D. Kavanaugh, B. Zimmerberg, & S. Fein (Eds.), Person behavior: Interdisciplinary perspectives, 1996, pp. 139-170. Mahwah, NJ: Lawrence Erlbaum.

[6] A. Campbell, "Staying alive: Evolution, culture and intra-female aggression," Behav. Brain. Sci., vol. 22, pp. 203-252, 1999.

[7] C. Kirschbaum, D. H. Hellhammer, "Salivary cortisol in psychoneuroendocrine research: Recent developments and applications." psychoneuroendocrino, vol.19, pp. 313–333, 1994.

[8] O. Houstis, S. Kiliaridis, "Gender and age differences in facial expressions," Eur. J. Orthod., vol. 31(5), pp. 459-466, 2009.

[9] C. Sforza, A. Mapelli, D. Galante, S. Moriconi, T. Ibba, L. Ferraro,V. Ferrario, "The effect of age and sex on facial mimicry: a threedimensional study in healthy adults,", Int. J. Oral. Maxillofac. Surg., vol. 39(10), pp.990-999, 2010.

[10] D. McDuff, E. Kodra, R. Kaliouby, M. LaFrance, "A large-scale analysis of sex differences in facial expressions," PLoS. One., vol. 12(4),2107.