DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

PROJECT WORK PHASE II

Progress Report

Batch No: 7 **Date: 31-05-2021**

Project Title: EMOTINET-A FACIAL EXPRESSION RECOGNIZER

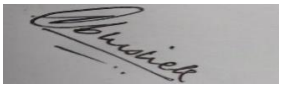
Team Members:

USN	NAME
1MV17CS002	ABHISHEK BHALOTIA
1MV17CS003	ADIYA RAJNEESH SINGH
1MV17CS061	NAMAN SAXENA
IMV16CS039	KALYAN KUMAR

Guide	**Dr. SUSHILA SHINDAL**

TABLE OF CONTENTS

Signature of Student

ABHISHEK BHALOTIA 

ADIYA RAJNEESH SINGH 

NAMAN SAXENA 

KALYAN KUMAR 

I. INTRODUCTION

According to reports from various web sources, the most widely used mode of communication used by humans is Facial expression. Facial expression recognition is an evolving technology in the field of human-computer interaction. Facial expression recognition has its branches spread across various applications such as virtual reality, webinar technologies, online surveys and many other fields. Even though high advancements have been witnessed in this field, there are several diplomacies that exist. The traditional feature extraction methods have slower response and lack in performance. The traditional methods have high latency or delay in their response. Through these traditional methods, it is extremely difficult to extract the required features effectively and hence, is too hazardous to utilize for real time applications. In order to provide a panacea to such issues, a facial recognition method is proposed using CNN. This model thus, can be used to solve the above stated problems or difficulties. The development of the facial expression recognition is done in Keras. A six layered CNN is developed to build and train the model. Each layer is defined with certain training techniques to enable faster and efficient feature extraction. The trained model is deployed to a web interface using Flask app. The developed model is applied for real time videos and images and its accuracy is analyzed.

The objectives of this project work are as follows:-

- To explore the dataset FER-2013.
- To generate training and validation batches.
- To create a Convolutional Neural Network (CNN) model.
- To train and evaluate model.
- To represent the model as JSON string.
- To create a Flask app to serve predictions.
- To design an HTML template for the Flask app.
- Use the model to recognize facial expressions in real time and analyze its accuracy.

II. LITERATURE SURVEY

a. Facial Expression Recognition with Faster R-CNN.

In order to avoid the complex explicit feature extraction process and the problem of low-level data operation involved in traditional facial expression recognition, we proposed a method of Faster R-CNN (Faster Regions with Convolutional Neural Network Features) for facial expression recognition in this paper. Firstly, the facial expression image is normalized and the implicit features are extracted by using the trainable convolution kernel. Then, the maximum pooling is used to reduce the dimensions of the extracted implicit features. After that, RPNs (Region Proposal Networks) is used to generate high-quality region proposals, which are used by Faster R-CNN for detection. Finally, the Softmax classifier and regression layer is used to classify the facial expressions and predict boundary box of the test sample, respectively. The dataset is provided by Chinese Linguistic Data Consortium (CLDC), which is composed of multimodal emotional audio and video data. Experimental results show the performance and the generalization ability of the Faster R-CNN for facial expression recognition. The value of the mAP is around 0.82.

b. Real Time Facial Expression Recognition in Video using Support Vector Machines.

Enabling computer systems to recognize facial expressions and infer emotions from them in real time presents a challenging research topic. In this paper, we present a real time approach to emotion recognition through facial expression in live video. We employ an automatic facial feature tracker to perform face localization and feature extraction. The facial feature displacements in the video stream are used as input to a Support Vector Machine classifier. We evaluate our method in terms of recognition accuracy for a variety of interaction and classification scenarios. Our person-dependent and person-independent experiments demonstrate the effectiveness of a support vector machine and feature tracking approach to fully automatic, unobtrusive expression recognition in live video. We conclude by discussing the relevance of our work to affective and intelligent man-machine interfaces and exploring further improvements.

c. **Real-time Emotion Recognition from Facial Expressions.**

We built several models capable of recognizing emotions from facial expressions. Using the FER-2013 dataset of non-posed grayscale images, we achieve 47.8% accuracy using an SVM and 66.5% using a CNN; on the CK+ dataset, we achieve 99.5% accuracy. We then built a real-time system to detect faces from a video feed and continuously classify them using our model, demonstrating the ability to transfer skills learned on the static datasets.

d. **Baseline CNN structure analysis for facial expression recognition.**

We present a baseline convolutional neural net-work (CNN) structure and image pre-processing methodology to improve facial expression recognition algorithm using CNN. To analyse the most efficient network structure, we investigated four network structures that are known to show good performance in facial expression recognition. Moreover, we also investigated the effect of input image pre-processing methods. Five types of data input (raw, histogram equalization, isotropic smoothing, diffusion-based normalization, difference of Gaussian) were tested, and the accuracy was compared. We trained 20 different CNN models (4 networks x 5 data input types) and verified the performance of each network with test images from five different databases. The experiment result showed that a three-layer structure consisting of a simple convolutional and a max pooling layer with histogram equalization image input was the most efficient. We describe the detailed training procedure and analyse the result of the test accuracy based on considerable observation.

e. **Understanding of a convolutional neural network.**

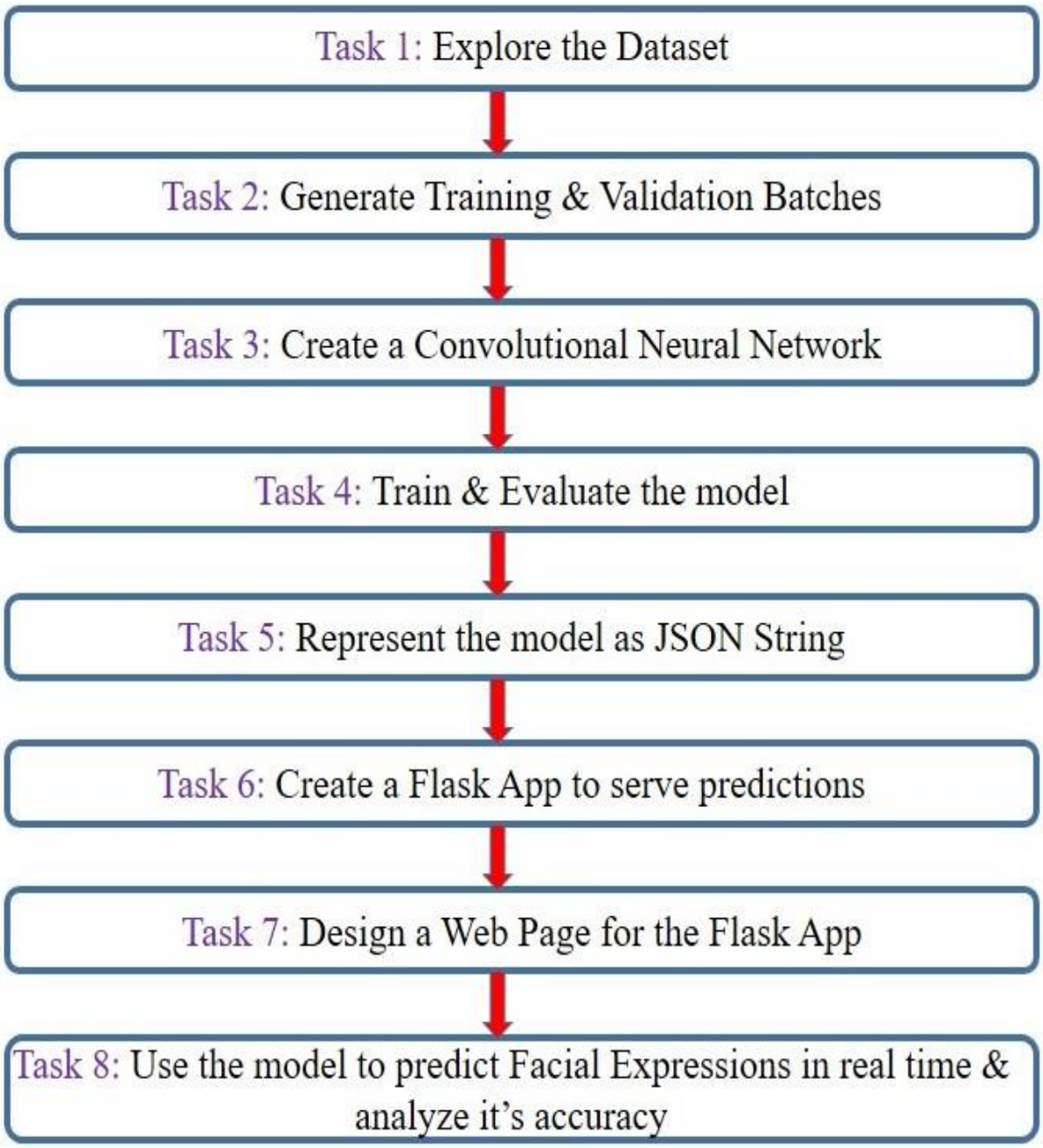
The term Deep Learning or Deep Neural Network refers to Artificial Neural Networks (ANN) with multi layers. Over the last few decades, it has been considered to be one of the most powerful tools, and has become very popular in the literature as it is able to handle a huge amount of data. The interest in having deeper hidden layers has recently begun to surpass classical methods performance in different fields; especially in pattern recognition. One of the most popular deep neural networks is the Convolutional Neural Network (CNN). It take this name from mathematical linear operation between matrixes called convolution. CNN have multiple layers; including convolutional layer, non-linearity layer, pooling layer and fully-connected layer. The convolutional and fully-connected layers have

parameters but pooling and non-linearity layers don't have parameters. The CNN has an excellent performance in machine learning problems. Especially the applications that deal with image data, such as largest image classification data set (Image Net), computer vision, and in natural language processing (NLP) and the results achieved were very amazing. In this paper we will explain and define all the elements and important issues related to CNN, and how these elements work. In addition, we will also state the parameters that effect CNN efficiency. This paper assumes that the readers have adequate knowledge about both machine learning and artificial neural network.

III. PLAN OF ACTION

Task	Due date	Outcome	Challenges faced/Comments
Publication of the Review paper with respect to phase I	20th December 2020	Finished successfully	Referring of various base papers, checking of plagiarism and grammatical errors in the paper.
Presentation of Phase I	20th January 2021	Finished successfully	Presented in front of the project guide through Google meet.
Beginning of the Phase II work	1st February 2021	In Progress	Perform various tasks in order to complete the project work.
Publication of the final paper.	Late June 2021 to July 2021	To be done	Read more 4-5 latest papers and decide the publisher where the paper should be published.
Presentation of final working project.	Late June 2021	To be done	----------
Submission of the presentation, project report, paper to the department.	Late June 2021	To be done	----------

IV. METHODOLOGY



V. CONCLUSION

In recent years there has been a growing interest in improving all aspects of the interaction between humans and computers. This emerging field has been a research interest for scientists from several different scholastic tracks, i.e., computer science, engineering, psychology, and neuroscience. These studies focus not only on improving computer interfaces, but also on improving the actions the computer takes based on feedback from the user. Feedback from the user has traditionally been given through the keyboard and mouse. Other devices have also been developed for more application specific interfaces, such as joysticks, trackballs, data gloves, and touch screens. The rapid advance of technology in recent years has made computers cheaper and more powerful, and has made the use of microphones and PC-cameras affordable and easily available. The microphones and cameras enable the computer to "see" and "hear," and to use this information to act.

Human beings possess and express emotions in everyday interactions with others. Emotions are often reflected on the face, in hand and body gestures, and in the voice, to express our feelings or likings. While a precise, generally agreed upon definition of emotion does not exist, it is undeniable that emotions are an integral part of our existence. Facial expressions and vocal emotions are commonly used in everyday human-to-human communication, as one smiles to show greeting, frowns when confused, or raises one's voice when enraged.

Psychologists and engineers alike have tried to analyze facial expressions in an attempt to understand and categorize these expressions. This knowledge can be for example used to teach computers to recognize human emotions from video images acquired from built-in cameras. For example, knowing the user's emotions, the computer can become a more effective tutor. Synthetic speech with emotions in the voice would sound more pleasing than a monotonous voice. Computer "agents" could learn the user's preferences through the users' emotions. Another application is to help the human users monitor their stress level. In clinical settings, recognizing a person's inability to express certain facial expressions may help diagnose early psychological disorders.

REFERENCES

[1] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186.

[2] Yamashita, R., Nishio, M., Do, R.K.G. et al. Convolutional neural networks: an overview and application in radiology. Insights Imaging 9, 611–629 (2018).

[3] Deepti, Archana, Dr. Jagathy; "Facial expression recognition using ANN", IOSR Journal of Computer Engineering 2013.

[4] Ms.Aswathy.R "A Literature review on Facial Expression Recognition Techniques," IOSR Journal of Computer Engineering (IOSR-JCE) 2013.

[5] Real Time Facial Expression Recognition in Video using Support Vector Machines (Philipp Michel Computer Laboratory University of Cambridge CB3 0FD, United Kingdom), (Rana El Kaliouby Computer Laboratory University of Cambridge, Cambridge CB3 0FD, United Kingdom).

[6] O'Shea K, Nash R. 2015.An introduction to convolutional neural networks. ArXiv: 1511.08458.