International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

EMOTION DETECTION AND ANALYSIS

USING CONVOLUTION NEURAL NETWORK

SREELAKSHMI S[1], Y ANUPAMA[2], UJWAL S[3], RAHUL GIREESH[4], SARANYA R[5]

[1]Student, Dept. of Electronics and Communication Engineering, NSS College of Engineering, Kerala, India
[2]Student, Dept. of Electronics and Communication Engineering, NSS College of Engineering, Kerala, India
[3]Student, Dept. of Electronics and Communication Engineering, NSS College of Engineering, Kerala, India
[4]Student, Dept. of Electronics and Communication Engineering, NSS College of Engineering, Kerala, India
[5]Assistant Professor, Dept. of Electronics and Communication Engineering, NSS College of Engineering, Kerala, India

---

Abstract - *Electroencephalogram (EEG) signal processing finds diverse applications including disease identification and monitoring, sleep studies, brain computer interface, biofeedback machines. Emotion detection from EEG signal is a leap frog in the field of medical science and engineering technology as it could help detect the emotion of facially paralyzed patients, infants, human emotional stress management and early detection of diseases. The classification of emotions using novel classification methods such as Convolution Neural Network (CNN) would enable the development of more accurate emotion detection systems. This work implements an efficient emotion detection and analysis software that will work on DEAP dataset, which is a three dimensional recording of EEG and peripheral physiological signals from 32 participants. In order to improve the accuracy of emotional recognition by end-to-end automatic learning of emotional features in spatial and temporal dimensions of Electroencephalogram (EEG), an EEG emotional feature learning and classification method using Convolution Neural Network (CNN) was proposed based on temporal features, frequential features, and their combinations of EEG signals in DEAP dataset. The extracted features will be given to the CNN classifier and the accuracy of the result is compared with existing classifiers.*

Key Words: **CNN, EEG, DEAP Dataset, Emotion detection**

1. INTRODUCTION

Emotions are vital in human decision handling, interaction and cognitive process. As technologies are advancing, there are growing opportunities for automatic emotion recognition systems. There are successful research breakthroughs on emotion recognition using text, speech, facial expressions or gestures as stimuli. However one in every of the new and exciting directions this research is EEG-based technologies for automatic emotion recognition, because it becomes less intrusive and cheaper, resulting in pervasive adoption in healthcare applications. In this paper we target on classifying user emotions from Electroencephalogram (EEG) signals, using different neural network models and advanced techniques. Particularly Deep Neural Networks and Convolutional Neural Networks, using advanced machine learning techniques like Dropout technique are used for emotion classification. Neural network could be a machine that's designed to model the way our brain performs a selected task, where the key concepts of brain as a posh, non-linear and parallel computer are imitates , and possess the flexibility to model and estimate complex functions looking on multitude of things. Moreover recent developments in machine learning have shown neural networks to supply prime accuracy in varied tasks like Text and Sentiment Analysis, Image recognition, and Speech analysis. Recently, the affective EEG benchmark database DEAP was published, which presents multimodal data set for the analysis of human affective states.

The Electroencephalogram (EEG) and peripheral physiological signals of 32 participants were recorded as each watched 40 one-minute long excerpts of music videos. Participants rated each video in terms of the amount of arousal, valence, like/dislike, dominance, and familiarity .A 32 EEG channels Bio-semi Active Two device was accustomed record the EEG signals when the themes were exposed to the videos. Aside from the EEG recordings, channels also recorded some physiological signals like temperatures and respiration etc. Methods and results were presented for single-trial classification of arousal, valence, and like/dislike ratings using the modalities of EEG, peripheral physiological signals, and multimedia content analysis. Automatic classification of human emotion using EEG signals is researched well by various scholars. However within the release of DEAP data, research academia finds a homogenous dataset to effectively measure and compare accuracies for various classification algorithms.

The proposed model uses a Convolution Neural Network model designed to classify images effectively. The model uses two Convolution layers with Tan Hyperbolic as the Activator, followed by Max Pooling at the output. The resulting output is referred after applying dropout, before being fed to fully connected neural layer which feeds the output to the Convolution neural layer of classification

classes size, using Softplus as activator. The Convolution model is standard for advanced MNIST or CEAP Image classification. The Model achieves appreciable accuracies of 75.58% and 73.28% for Valence and Arousal respectively for classification on two classes (high and low); 58% and 54% for Valence and Arousal respectively for classification on 3 classes (high, normal and low). However it's our Convolution Neural Model which surpasses our Deep Model's accuracy by providing 81.41% and 73.36% for 2 class classification and 66.79% and 57.58% for 3 classes on Valence and Arousal respectively. Both these models provide state of-the-art classification accuracy reported on DEAP dataset, substantially improving classifications by previous research which struggled to achieve 75% and 55% on classification on two and three classes respectively. Furthermore, the tactic of representing EEG data in a very similar manner to that of an image and consequently using the representation as images to feed the Convolution Neural Model, exploiting the accuracy of CNNs on image classification.

2. RELATED WORKS

Emotion is a psycho-physiological process triggered by Conscious and/or unconscious perception of an object or situation and this emotion is typically associated with mood, temperament, personality, disposition, and motivation. Apart from the above mentioned factors emotion has a forehand on human communication as well. It can be expressed either verbally through emotional vocabulary or by expressing nonverbal cues like intonation of voice, facial expressions, and gestures [1]. The recent public release of DEAP dataset provides a far needed impetus to the growing community of HCI researchers in emotion recognition. Before DEAP, most of the studies on emotion assessment had focused on the analysis of facial expressions and speech to determine someone's spirit. However, physiological signals are known to include emotional information which can be used for emotion assessment, but they have received less attention [3]. The database explores the likelihood of classifying emotion dimensions induced by showing music videos to different users, using the signals originating from the central nervous system (CNS) and thus the peripheral nervous system (PNS).

DEAP uses Russells valence-arousal scale, widely utilized in research on affect, to quantitatively describe emotions. During this scale, each spirit is placed on a 2D plane with arousal and valence because the horizontal and vertical axes. Arousal can range from inactive (e.g. uninterested, bored) to active (e.g., alert, excited), whereas valence ranges from unpleasant (e.g. sad, stressed) to pleasant (e.g., happy, elated) as given in fig 1. It contains EEG and peripheral physiological signals recorded employing a Biosemi Active Two system at a rate of 512 Hz using 32 active AgCl electrodes (placed per the international 10-20 system) [6]. DEAP has sufficient participants in publicly available databases for analysis of spontaneous emotions from physiological signals. Moreover, it is the sole database that uses music -videos as emotional stimuli. Since the discharge of DEAP dataset, multiple researchers are using it for emotion recognition. (Liuand Sourina (2014)) research, explores real-time Electroencephalogram (EEG)-based emotion recognition algorithm using Higuchi Fractal Dimension (FD) Spectrum. They recognize EEG as a nonlinear and multi-fractal signal, hence its FD spectrum can provides an improved understanding of the nonlinear property of EEG using Support Vector Machines as a classifier. The approach was tested on both DEAP and their own databases. DEAP database gave an accuracy of 53.7% on 8 emotions in subject dependent classification.

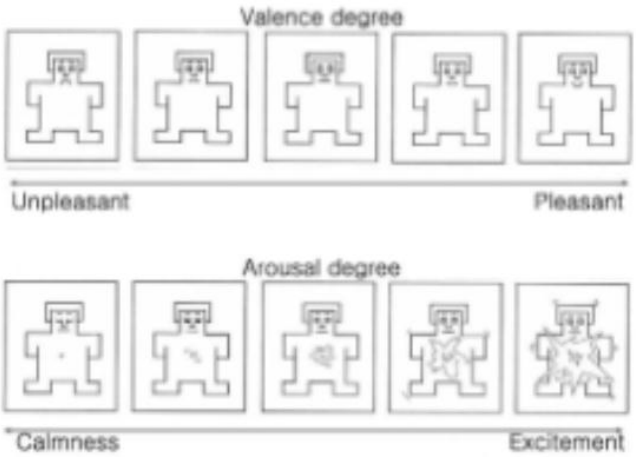


Figure 1: Degree of valence arousal model [4]

In (Srivastava et al (2014)) research, depicts models for Classification of DEAP's EEG data to different energy bands using wavelet transform and neural networks. They divide EEG signal into different bands using discrete wavelet transformation with db8 wavelet function for processing. Statistical and energy based features are extracted from the bands, supported the features emotions are classified with feed forward neural network with weight optimized algorithm like PSO [2]. The research builds upon the works of (Chung and Yoon(2012)) which focuses mainly on classification of DEAP data into classes of Valence and Arousal using statistical and shallow learning methods like Bayesian Classification. Their simple classification methods provide a starting baseline for our study to match results. They classify the user data into two(high and low) and three(high, normal, low) classes for both Valence and Arousal. They achieve 66.6% and 66.4% accuracy for two classes, and 53.4% and 51.0% for three classes, on Valence and Arousal.

Figure 2: Valence- Arousal model [3]

The work of (Candra et al (2015)) investigates the how the window size effects the classification of DEAPs, EEG data using wavelet entropy and SVMs. They conclude that a very wide window may result in information overload which causes the reduction of feature extraction accuracy. Similarly the information about emotion won't be adequately extracted if the time window is just too short. Discrete wavelet transform (DWT) coefficient is used for extracting time-frequency domain features in EEG signals. The investigation revealed that arousal is classified up to 65.33% accuracy using the window length of 310 seconds; while valence is classified upto 65.13% accuracy using the window length of 312 seconds. (Sohaibetal(2013)) provide a concise evaluation for various classifiers for Emotion Recognition. However instead of the DEAP data, their test was done using another EEG dataset of 20 subjects which was subjected to photographs from International Affective Picture System (IAPS). They evaluated the classification for K-Nearest Neighbor (KNN), Bayesian Network (BN), Artificial Neural Network (ANN) and Support Vector Machine (SVM).Their results showed that it's difficult to train a classifier to be accurate over big datasets but KNN and SVM with the proposed features were reasonably accurate over smaller datasets identifying the emotional states with an accuracy up to 77.78% [10].

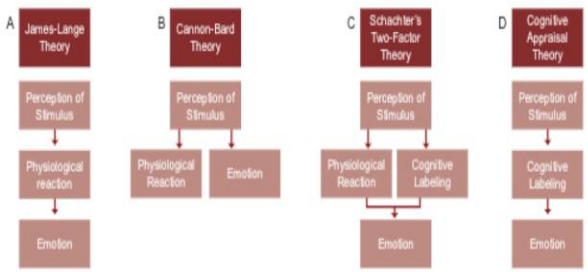


Figure 3: Emotion theories with A:James-Lang Theory, B: Cannon-Bard Theory, C:Schachter-Singer Theory, D:Cognitive Appraisal Theory [3]

The research provides the best classification accuracy on the DEAP data for both Arousal and Valence, which we hopefully improve upon. Their innovative technique is based on three steps : Firstly, the EEG signal is represented as a sequence of overlapping segments and extract feature vectors on the segment level; Secondly segment level features are transformed into the response level features using projections supported a totally unique non-parametric nearest neighbor model; Thirdly the obtained results are classified [9]. Preprocessing was done using KPCA dimensionality reduction. The information was deployed using classification algorithm such as Naive Bayes Nearest Neighbor, RBF SVMs and Nearest Neighbor Voting. Leave-one response-out cross validation scheme was used for the validation of individual and mean accuracies.

3. DATA ACQUSITION, PROCESSING AND FEATURE EXTRACTION

3.1 Deap Dataset

The DEAP dataset consists of the ratings from an online self-assessment, where 120 1 minute extracts of music videos were each rated by 14-16 volunteers based on valence, arousal and dominance and the participant ratings, face video and physiological recordings of an experiment where 32 volunteers watched a subset of 40 of the above music videos. EEG and physiological signals were recorded. Participants rated the videos as above. For 22 participants their frontal face video was also recorded. The official dataset contains all individual ratings from the online self-assessment, YouTube links of the music videos used, all ratings participants gave to the videos, the answers participants gave to the questionnaire before the experiment, the frontal face video recordings from the experiment for participants 1-22 and the original unprocessed physiological data recordings from the experiment in BioSemi.bdf format. However for the experiment we use the preprocessed (data down sampled to 128Hz, EOG removal, filtering, segmenting etc.) physiological data recordings from the DEAP experiment in Matlab and Python (numpy) format. This format is particularly useful for testing classification or regression methods without the difficulty of processing all the data. For each of the 32 participants there exist 2 arrays illustrated in Table 1. The dataset consists of 40 experiments for each of the 32 participants. The labels array contains the valence, arousal, dominance and liking ratings.

3.2 Processing and Feature Extraction

The data array contains 8064 EEG signal data from 40 different channels for each of the 40 experiments for each of the 32 participants. As one can see, for each experiment we have a massive 322560 readings to train the classification algorithm. To allow the neural models so it could effectively

and speedily train on such massive data, it is to be proceeded to reduce dimensionality of data. The 8064 readings per channel were divided into 10 batches of approximately 807 readings each. For each batch the mean, median, maximum, minimum, standard deviation, variance, range, skewness and kurtosis values for the 807 readings were extracted. Then for each of the 10 batches of a single channel 9 value mentioned above were extracted. 90 values as processed dataset is attained. Then further add the net median, minimum, mean and standard deviation, maximum, variance, range, skewness and kurtosis values for all the 8064 readings along with the experiment and participant number to the dataset, making it up to 101 values / channel. As mentioned in the work of (Candra et al (2015)), the optimal sliding window size of 310 and 3-12 seconds was ideal for classification of Valence and Arousal respectively.

8064 readings represent the EEG values recorded over the duration of 1 minute of the participant viewing the video; classifying them into 10 batches gives us a brief of emotions for a 6 second range. Moreover statistical methods to reduce the dimensionality of the EEG dataset mapping its feature probability density function to a Gaussian distribution is obtained and then effectively catching it using statistical features like mean, variance range etc (Gupta and Gupta (2009)). (Jahankhani et al (2007)) effectively demonstrate this method using maximum, minimum, mean and standard deviation of wavelet coefficients for signal classification. Then for each of the 10 batches of a single channel we extract 9 statistical values and the net mean of these features using Principal component Analysis. PCA speeds up model training time and in most cases improves accuracy. It is the most widely used method for pattern recognition and feature extraction. PCA is used when there are a large number of variables and there occurs some redundancy in the variables. Due to this redundancy, it is possible to reduce the observed variables into a smaller number of principal components that will account for most of the variance in the observed variables. When Principal Component Analysis is used with Neural Network, initially the redundant data in the dataset is eliminated and the data obtained is trained using Neural Network and Projects higher dimensional data to lower dimensional data

In order to reduce the volume of their EEG data samples were partitioned into 16 windows of 256 time points each. A similar statistical dimensionality reduction was followed for our dataset as well. Finally the 322560 readings per experiment were reduced to 4040 values (101 reduced readings * 40 channels). These values per experiment form our initial processed dataset, where used to train the neural model. As mentioned before the use of leave-one response-out cross validation is done. This implies that for participant 1, the model is trained using readings for participant 2 to 32 and record classification accuracy for participant 1. For participant 2, the model was trained with the same architecture but this time it is trained using readings for participant 1 and 3 to 32 and record the classification accuracy for participant 2 and so on. This allows us to train the model for 1240 experiments (31 participants * 40 experiments) and predict for the 40 experiments for each of the subjects one after another. The labels data are iterated and for each of Valence and Arousal we extract one hot encoding based outputs for classification in both two classes (ratings divided as greater than and less than 5), and three classes (ratings divided as greater than 6, between 4 and 6, smaller than 4).

3.3 Model Training and Classification

The extracted features now reach the training model. We convert this image into a 2D array image. This 2D array image will undergo multiple rounds of iterations, each such iteration is known as a step. In this project the model is trained in valence and arousal domain. The valence model is concerned about the positive and negative emotions and Arousal model deals with the active and passive state of emotion. As the iteration progresses, there will be more ideal separation between multiple emotions. Convolutional neural network (CNN) models are one of the effective solutions for Image classification tasks. For the model, we try and convert our DEAP data into 2D image format so the CNN model can learn to classify them effectively. Here for each experiment there are 40 channels with 101 readings each. So the data for each experiment has a 2D array image of 40 * 101 size. First Convolution layer takes this 2D array as input and the Convolution operation uses 100 initial convolution filters and a convolution kernel of three rows and three columns. The first Convolution layer uses 'Tan Hyperbolic' as the Activation function for Valence Classification model, and 'Relu' or Rectilinear units as Activation for Arousal model. In this experiment it is realized that the choice the Activation functions for the first layer are of cardinal importance, as some functions (like sigmoid, softmax) might not be able to activate neurons of later layers consistently, making the model defective.

The next layer is another Convolution Neural Layer which again with 100 filters and 3*3 size kernel. This layer uses 'Tan Hyperbolic' as the Activation function for both Valence and Arousal classification. The next layer is a MaxPooling layer, and the pooling is traditional 2 dimensional max pooling over 2x2 blocks. The use of Dropout on the outputs of MaxPooling layer, with a Dropout probability of 0.25, to form a Flat 1 dimensional layer is established. The output of the layer is fed into a Fully connected Dense neural layer that gives an output with dimensionality of 128. We use 'Tan Hyperbolic' as our activation function again use Dropout with 0.5 probability on the Outputs of Dense layer. Finally the final Fully Connected Dense neural layer has an output dimensionality of two or three, depending on the number of output classes. The final Dense layer uses 'Softplus' as its activation function. The model uses the Categorical Cross Entropy as the loss function and Stochastic gradient descent (SGD) as the optimizer with a learning rate of 0.00001 for

Valence and 0.001 for Arousal and gradient momentum of 0.9. For 4749 experiments we use 250 epochs and train the model using batches of 50 experiments each.

The Convolution layer is the core building block of a Convolution Network. The output of neurons that are connected to local regions in the input, are calculated by performing the dot product between their weights and the region they are connected to the input volume. CNN works as follows, during the forward pass, we slide (more precisely, convolve) each filter across the width and height of the input volume, producing a 2-dimensional activation map of that filter. As the filter is slided across the input, the dot product between the entries of the filter and the input are computed. Once a filter has been glazed over the complete input, we find the single most important feature using max-over-time pooling operation (POOL Layer). This allows us to correctly identify one feature for each filter. The model repeats this for each filter in the image, to obtain best features for an experiment, in each convolution. Supporting these features for all filters along the depth dimension forms the full output volume.

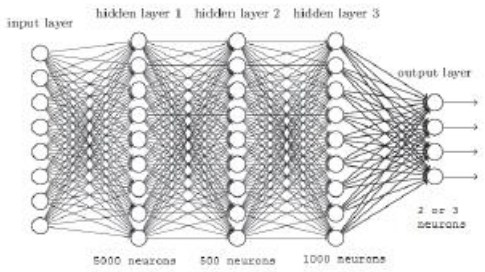


Figure 4: Convolutional Neural Network implementation using three hidden layers [7]

Thus, every output volume can also be represented as an output of a neuron that covers only a small region in the input and shares parameters with neurons in the same activation map. The Max Pooling layer's function is to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network, and hence to also control over fitting. Using the MAX operation the Pooling Layer independently operates on the input and it is spatially resized. Neurons in the fully connected layers have complete connections with the previous layers to all activation functions. Hence their activations can be computed with matrix multiplication followed by a bias offset.

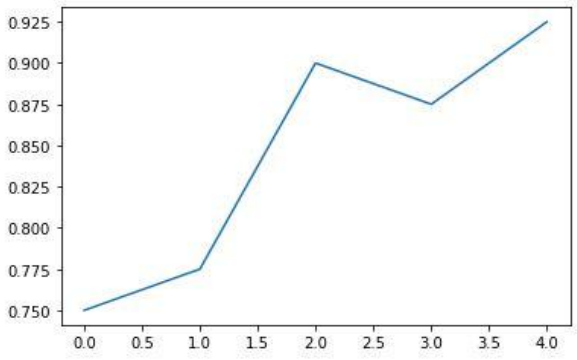
4. RESULT

In the result, we got the feature extracted output. Here it is considered that the signal of a single participant from a single channel. The graph obtained after the feature extraction is in the form of EEG signal which depicts the transformation of raw or unprocessed EEG signal to feature extracted EEG signal. This feature extracted output is used for training. PCA optimize the accuracy of the model by 6% on average.

During experimentation the highest explained variance by a single variable was found to be 43% and the lowest explained variance was found to be 0.04%.

After feature extraction, Model training is done. In model training first of all the extracted features for all 32 participants are loaded into the program for model training. In machine learning algorithm the valence- arousal model is used for emotion training and Classification. Now each model is trained individually using the extracted features. While training the valence model the statistical distribution of values on the valence axis is learned by the model. Its learning accuracy is found to be around 97%. This will be saved as a file and called at the time of classification and prediction.

Similarly the features are fed to the arousal model for learning the statistical distribution of parameters along the arousal axis. This is in order to classify emotions based on their intensity. For example low arousal means a calm state & high arousal means excited state. This is also saved as a file. The accuracy is around 99%. It is believe that the ideal behavior of arousal model (100%) is due to some error in the rapid test case which was quiet unpredictable. Graph 1 depicts the accuracy of the valence model during our experiment.



Graph 1: Valence model accuracy after training

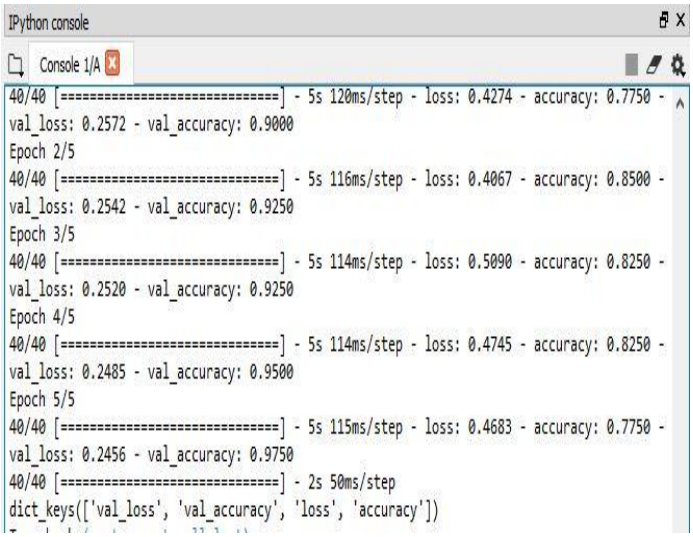


Figure 5: Accuracy of valence model- 97%

Classification is done using CNN which depicts the most dominant and less expressed emotion in multiple epochs.

The first convolution layer learns the basic features. The second layer learns more complex features. The maxpool layer is used to reduce over fitting issue due to training. The flat 1D layer enhances system performance. Dropout function randomly rejects some neurons there by reducing some data in order to prevent over fitting. Dropout of 25% means 25 percent of data is dropped down.

Activation functions are applied on arousal and valence models to make them create a desired output. Relu, Softmax, Tan Hyperbolic are some of the activation functions that have been applied in the CNN layers.

The below given figure 7 is a random case which predicts the emotion Happy. This is the output from 40 channels of one participant. The dominant emotion here is happiness. Graphical user interface (GUI) is also developed for the front end users for its easy manipulation.

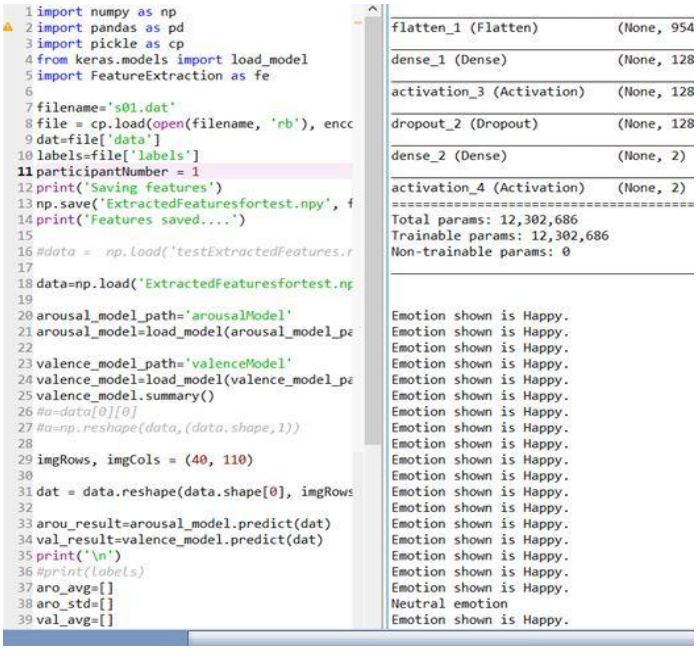


Figure 6: Illustration of layers in CNN and happy emotion prediction

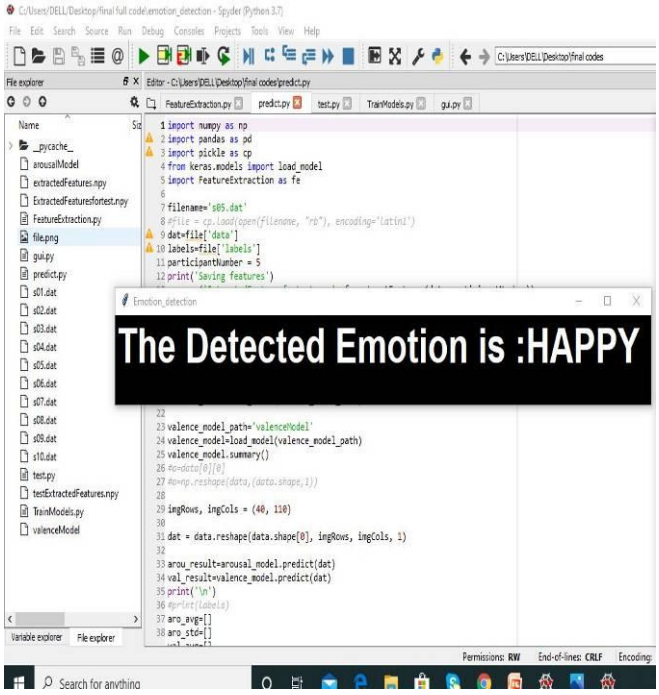


Figure 7: Happy emotion predicted with GUI

5. CONCLUSION AND FUTURE SCOPE

Even though there will be a promising result from our study, there is still room for the improvement to get even better result. Such improvement can be achieved by adding more subjects which will result in increase in dataset size. Other technique like Peak detection, Thresholding can be used for EEG data processing. EEG signals can also be isolated into

separate frequency band of Alpha and Beta to check if that gives some interesting result. Most emotion recognition systems are used to classify the specific emotion states, such as happiness, sadness. Studies on the evolution process of the emotion state are limited, which is not conducive to learning a person's emotion changing process. It is expected that the combinations of physiological signals, including EEG, ECG, GSR, RSP, EMG, HR, will lead to significant improvements in emotion recognition in the coming decade, and that this recognition will be critical to impart machines the intelligence to improve people's life.

We can apply these computational models to data acquired with wearable devices, for the recognition of emotion from physiological signals. In addition, employing a robust and novel feature extraction, feature selection and classification techniques would help in developing a user independent emotion recognition system with higher classification accuracy.

Furthermore the entire system performance is validated on the grounds of accuracy, learning rate, sensitivity and specificity.

Comparison of CNN's performance against existing machine learning algorithm such as Kth Nearest Neighbour (KNN), Support Vector Machine (SVM), Dense Neural Network (DNN) enhances the reach of the developed project which is given below.

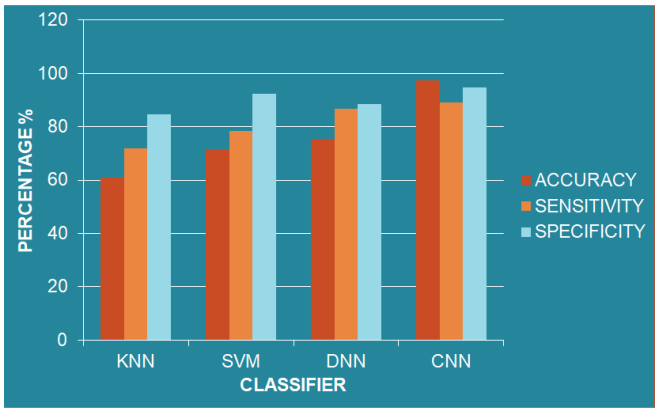


Table 1: Performance Comparison of our CNN model with respect to KNN, SVM and DNN machine learning models

REFERENCES

[1] Tripathi, Samarth, et al. "Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset." Twenty-Ninth IAAI Conference. 2017.
[2] Chen, J. X., et al. "Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks." IEEE Access 7 (2019): 44317-44328.
[3] Koelstra, Sander, et al. "Deap: A database for emotion analysis; using physiological signals." IEEE transactions on affective computing 3.1 (2011): 18-31.
[4] Horlings, R. "Emotion recognition using brain activity [dissertation].[Delft Netherlands]: Faculty of Electrical Engineering, Mathematics, and Computer Science." Delft University of Technology (2008).
[5] Bazgir, Omid, Zeynab Mohammadi, and Seyed Amir Hassan Habibi. "Emotion Recognition with Machine Learning Using EEG Signals." 2018 25th National and 3rd International Iranian Conference on Biomedical Engineering (ICBME). IEEE, 2018.
[6] Yang, Heekyung, Jongdae Han, and Kyungha Min. "A Multi-Column CNN Model for Emotion Recognition from EEG Signals." Sensors 19.21 (2019): 4736.
[7] Khosla, Sopan. "EmotionX-AR: CNN-DCNN autoencoder based emotion classifier." Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media. 2018.
[8] Zheng, Bong Siao, M. Murugappan, and Sazali Yaacob. "Human emotional stress assessment through Heart Rate Detection in a customized protocol experiment." 2012 IEEE Symposium on Industrial Electronics and Applications. IEEE, 2012.
[9] Mehmood, Raja Majid, Ruoyu Du, and Hyo Jong Lee. "Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors." Ieee Access 5 (2017): 14797-14806.
[10] Xu, Ruifeng, et al. "An ensemble approach for emotion cause detection with event extraction and multi-kernel svms." Tsinghua Science and Technology 22.6 (2017): 646-659.
[11] Rajabi, Zahra, Amarda Shehu, and Ozlem Uzuner. "A Multi-channel BiLSTM-CNN Model for Multilabel Emotion Classification of Informal Text." 2020 IEEE 14th International Conference on Semantic Computing (ICSC). IEEE, 2020.
[12] Oh, SeungJun, Jun-Young Lee, and Dong Keun Kim. "The Design of CNN Architectures for Optimal Six Basic Emotion Classification Using Multiple Physiological Signals." Sensors 20.3 (2020): 866..
[13] Oh, SeungJun, Jun-Young Lee, and Dong Keun Kim. "The Design of CNN Architectures for Optimal Six Basic Emotion Classification Using Multiple Physiological Signals." Sensors 20.3 (2020): 866.
[14] Jiang, Xinbei, and Tianhan Gao. "An EEG Emotion Classification System Based on One-Dimension Convolutional Neural Networks and Virtual Reality." International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing. Springer, Cham, 2020.