

# Hybrid CNN-RNN Model for Personality Detection

\*Princy Sathyadas, #Dr. Jayasudha J.S

\*PG Scholar, #Professor & Dean (Academic), SCT College of Engineering, Thiruvananthapuram, India , princysathyadas@gmail.com, jayasudhajs@gmail.com

**Abstract** - Personality detection based on texts from online social networks having more and more attentions recently. It has become a well known approach for data dispersion and social association. Character of a human is defined as the trademark units of practices, behaviors and emotional pattern etc. Personality detection models could be very useful in various domains like e-learning, information filtering, collaboration and e-commerce by a user interface that adapts the interaction according to user's personality and used for finding criminal cases, forensic, medical diagnosis and HR management. Here behavior of individuals can recognize dependent on the five factor model, The five-factor model of character (FFM) is a set of five wide characteristic measurements or domains, frequently known as the "Big Five" model that is, Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness.

This paper describes the state-of-art methods for personality detection such as CNN, LSTM and GRU, then these models are compared with hybrid CNN-RNN architecture. This model taking properties like feature extraction of CNN and learning long term dependency via RNN. In this hybrid model, variation of RNN such as CNN-LSTM and CNN-GRU is used that is generally known as Hybrid CNN-RNN model. The personality detection is done by using big five personality traits. Moreover, the comparison has been done among CNN, LSTM, GRU and hybrid CNN-RNN models. The hybrid models give better performance and accuracy than the existing models.

**Keywords** — Personality detection, Convolutional Neural Network, Long Short Term Memory, Gated Recurrent Unit

## I. INTRODUCTION

Nowadays, social media platforms are the largest and important mines of personal information, because they continuously record human's habits, interactions, interests in music, movies and shopping, etc. Our personality has an important part in our lives; it affects our life choices, well-being, health and numerous other preferences and everywhere. Therefore, automatic detection of an individual's behavior has numerous functional applications.

Personality is the most special human attribute, it also describes the distinctiveness of a person. Personality is one of the essential aspects, by which we can understand from different emotions of a person. However, in most of the cases, people respond with respect to their underlying behavior. It is possible to analyze large amounts of text samples then automatically identify personality types of authors and predict potential reactions and behaviors [1].

The specialty of the human is to understand others behavior based on the observation of their everyday activities, emotions or feelings and behavior. Several researchers especially in the field of artificial intelligence, computational linguistics, natural language processing etc

are around the world has been attracted to work on this domain machine interaction, behavioral analytics, and machine learning.

Personality is generally defined in terms of the big five personality traits, the five factors may be easily remembered using the acronym 'OCEAN'.

The 5-factor model of personality is the theory that there are five variables or factors that can be used to describe the human personality. Those five factors collectively known as Big 5 model for personality. Find the personality of an individual by using these big five personality traits model. That is extensively standard and used model for character detection.

Personality detection is a type of Text classification problem. Text classification problem has been addressed in many real applications over the last few decades [3]. Most of the text classification process and classification frameworks can be constructed with following four phases: Feature extraction, dimension reductions, classifier determination and assessments. The exponential development in the quantity of complex datasets consistently requires more improvements in machine

learning methods to provide strong and exact data classification such as deep learning approaches.

The success of these deep learning algorithms relies on their capacity to model complex and nonlinear relationships within the data. A convolutional neural network (CNN) is a type of artificial neural network utilized in image, text recognition and processing that is specifically designed to process pixel data. CNN has their “neurons” arranged more likely to processing visual stimuli in humans and other animals. Convolutional Neural Networks have a unique structure than ordinary Neural Networks. Regular Neural Networks include number of hidden layers, so it transforms an input by putting it through a series of hidden layers [4]. In CNN, each layer is formed from a group of neurons, where each layer is fully connected to all or any neurons within the layer before. In the case of NLP tasks, i.e., while carried out to textual content in place of images, we have a one dimensional array that represent the textual content. The architecture of the CNN is modified to 1D convolutional-and pooling operations that is used for text processing.

Recurrent Neural Network (RNN) is a kind of neural network architecture that is addressed by the researchers for text mining and classification. RNN assigns more weights to the previous data points of the sequence. Therefore, this system may be a powerful method for text, string and sequential data classification. In RNN, the output can be propagated in both direction, which allows for better semantic analysis of the structures in the dataset. Gated Recurrent Unit (GRU) and Long Short Term Memory (LSTM) [2] are two gating mechanisms for RNN, which is an efficient method for classification.

A hybrid supervised learning approach known as CNN-RNN fusion network that computes five personality traits by relying on what an individual thoughts or emotion. The approach segments input sentence in tokens, then it learns word vector representations of corresponding input as embeddings that are then used to feed a supervised learner classifier. Then this approach will be compared with the state-of-art methods for personality detection like CNN, LATM and GRU.

The paper is organized as follows, Section I contains the introduction of personality detection and different methods for personality detection, Section II explains the Big Five model, Section III describes about datasets, Section IV explain different personality detection models, Section V describes proposed methodology and its results and discussion are include in Section VI. Finally, the models used for personality detection are summarized and concluded in section VII.

## II. BIG FIVE PERSONALITY TRAITS MODEL

In recent years, the “Big Five” model of personality is most widely accepted and used model for personality detection. Personality include different unique characteristics of an individual. Two independent research teams derive this big five model in early 1990’s, they were asking thousands of questions from hundreds of peoples and analyze the data with a statistical procedure known as ‘Factor analysis’. The big five model is also referred to as OCEAN model. This model is the blueprint for understanding human behavior and probably it is the most reliable and accurate personality traits model. In this model, the behavior of a person can be classified into five categories such as Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism. This is the mixture of different emotions and characteristics. And it is a popular model in the field of sentiment analysis, crime analysis, Fraud detection, etc. Figure.2.1 shows the big five personality traits model and the properties of ‘Big five’ personality traits model is defined as follows:

**O (Openness):** Is the person is curious, innovative, careful, wise, and resourceful.

**C (Conscientiousness):** Is the individual is efficient, responsible and persevering. Conscientious people are extraordinarily reliable, difficult people and well deliberate human beings.

**E (Extraversion):** Is the man or woman is lively, assertive, outgoing, amicable and friendly. They have drawn thought from social conditions.



Fig.2.1 Big five personality traits model

**A (Agreeableness):** Is the person being Compassionate, cooperative, beneficial and straightforward. Agreeableness person is generally optimistic and trusting of others.

**N (Neuroticism):** Is the individual is hectic, self-pitying, irritating, insecure, touchy. Neurotics are always being

moody, traumatic, and frightened and worried approximately bad feelings.

Each of the Big Five personality traits consists of some similarities and they are correlated with each other. In this paper, personality detection can be done based on the user input. From the user input, classify a particular person on the basis of big five personality traits model.

### III. DATASET

The most popular labeled dataset are available for the evaluation of personality detection task, such as, Essays and myPersonality datasets. Personality detection model is widely used for many researchers. So that they developed their own dataset by selecting the text from various resources like social media, and several data sources. However, the huge number of researchers used the following dataset and now it is becoming more popular dataset to test and develop new personality detection models. Here we use Essay dataset for personality detection.

#### Essay dataset

Essays is a large dataset that contain about 2400 random essays, which is a random thought for each author collected between 1997 and 2004 and labeled with personality classes such as cEXT, cNEU, cAGR, cCON and cOPN. Each essay is labeled as the percentage of each traits then researchers and students took the data for Big Five personality test [5].

### IV. PERSONALITY DETECTION MODELS

Machine learning algorithms process only the fixed input and we get the fixed size of output. Suppose we want to process a long sequence of information that depends on previous input or output states that can be done by using recurrent neural network. In recurrent neural network, we share the weights and output of one node to the next node recursively. It creates a loop in the neural network architecture which act as a memory state of the neuron this state allow the network to remember the previous information and process the long sequence of information. We can use some advanced algorithms which includes Convolutional neural network and some deep learning algorithms (Long short term memory (LSTM) and Gated recurrent unit (GRU)) for personality detection. The objective is to assess the overall performance of various deep learning algorithms for personality detection and find the excellent method.

#### A. NAÏVE BAYES CLASSIFIER

Naïve Bayes is one of the most popular algorithm for text classification process. It is simple and perform more complicated solutions. This theorem basically states that

“The probability of A given that B is true equals the probability of B given that A is true times the probability of A being true, divided by the probability of B being true”. To process a sentence, first discover the possibility of every phrase inside the sentences and predict the word count in that sentence and must convert them in to probabilities and multiply them out we get the prediction as classification. The major disadvantage of this method is assumption for classification is independent from the predicted features.

#### B. SUPPORT VECTOR MACHINES

Support vector machine is an algorithm that determines boundary between vectors that belong to a category and vectors that do not belong to it. This algorithm applied to any vectors that encoded in to corresponding statistics. Here vectors are numbers, which represent a set of points in space. The data can mapped into a significant space, and then SVM divides the space into two subspaces: one for the vectors which belong to the given category and one for the vectors, which do not belong to it. To apply the SVM algorithm to text classification problems and obtain very good results.

#### C. LONG SHORT TERM MEMORY

The Long Short-Term Memory (LSTM) unit was initially proposed by Hochreiter and Schmidhuber [7]. It is a unique kind of recurrent neural network. Unlike other recurrent neural network, the main feature of LSTM is that it incorporates gated unit for store information and it contain a memory cell that decide which previous information to be remembered. LSTM computes input gate, forget gate and output gate to manage this memory. LSTM has designed to avoid the long-term dependency problem [8]. Practically the default behavior of LSTM is that it remembers the information for a long period, so that it can be used for processing long sequence of information. This method is used for various tasks like handwriting recognition, machine translation, sentiment analysis etc.

An LSTM architecture is first used for personality detection. Our dataset includes a longer sequence of text, it can be successfully processed via a network that containing gated unit to hold the previous information. Figure.4.1 indicates the LSTM architecture for personality detection.

In the first step, we tokenize the data. This is not a layer for the LSTM network, but a mandatory step of converting our words into tokens (integers). Then given the tokens into embedded layer that converts our word tokens (integers) into embedded of specific size. Next the LSTM layer that defines the hidden state dimensions and number of layers. The output of the LSTM layer is given to the fully connected layer that maps output of LSTM layer to a desired output size.

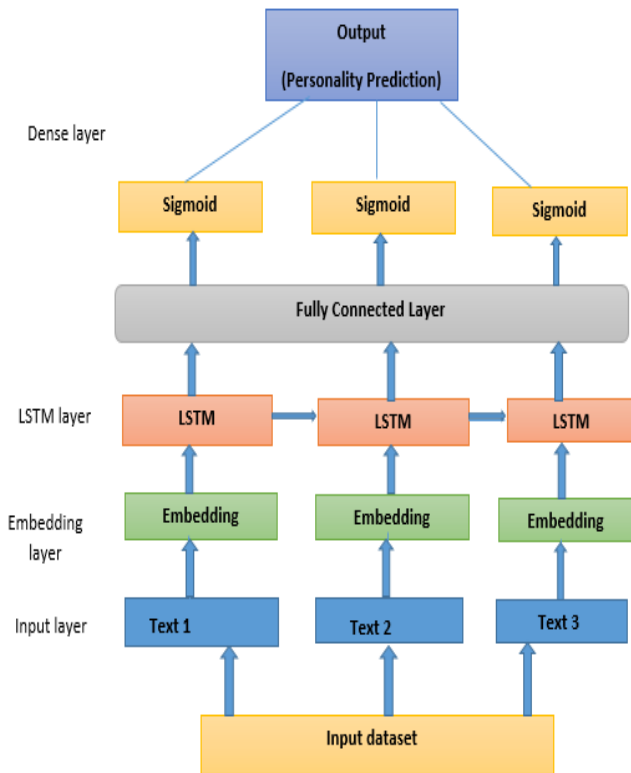


Fig.4.1 LSTM architecture for personality detection

The sigmoid activation function is applied for the output layer because of the gating mechanisms in LSTM. A parameter called return \_sequence having two values either TRUE or FALSE are specified in LSTM layer. If it is TRUE, each node in the LSTM layer provides its own output or it is FALSE, output provide only in the last time step. Finally, the LSTM makes classification and we get the result as the behavior of the person.

### B. GATED RECURRENT UNIT

The gated recurrent unit GRU is a relatively recent development proposed by Cho et al. just like the LSTM unit, GRU contains gating units that control the flow of information through the network, without having a separate memory cell. Unlike LSTM, Gated Recurrent Unit (GRU) having only two gates called update and reset gates, which control the flow of information through each hidden unit [9].

Number of gates and training parameters used in GRU are lesser than LSTM. Consequently, it use much less memory, and it executes and train faster than LSTM. The variant models carry out in addition to the unique GRU version while decreasing the computational price [10]. The operation of the GRU network is same as LSTM. In the first step, we have given the textual content information into embedding layer that converts our word tokens (integers) into embedding of specific size. Next the GRU layer that defines the hidden state dimensions and number of layers. The output of the GRU layer is given to the fully connected layer that maps output of GRU layer in to a preferred output

size. The sigmoid activation function is implemented for the output layer because of the gating mechanisms in GRU as in LSTM.

In GRU layer return sequence parameter, having two values either TRUE or FALSE is used. If it is TRUE, each node in the GRU layer provide its own output or it is FALSE, output provide only in the last time step. Figure 4.2 shows an example of GRU model.

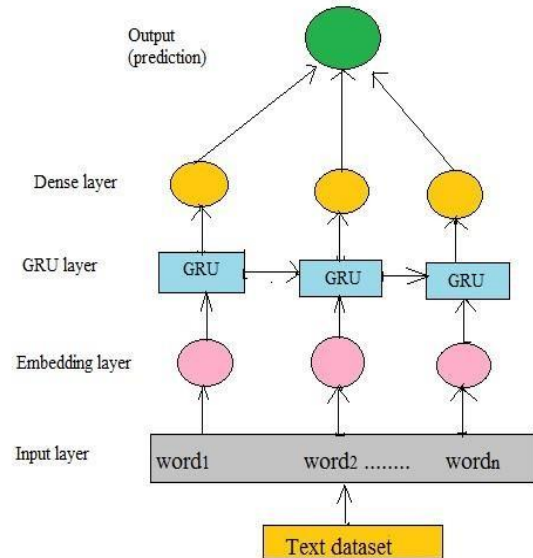


Fig.4.2 GRU architecture for Personality detection

### C. CONVOLUTIONAL NEURAL NETWORK

A convolutional neural network (CNN or ConvNet) is a sort of deep neural network. Convolutional Neural Networks (CNN) invented for image classification and it is used for text classification tasks and other traditional Natural language Processing (NLP) tasks. A convolutional neural network consists of a number of layers such as the input layer, the output layer, and multiple hidden layers. The hidden layers of a CNN usually include a sequence of convolutional layers that extract the feature from input contents. It has a RELU layer for applying an activation function and followed by additional layers like pooling layers, fully connected layers and normalization layers, these layers are known as hidden layers in CNN because their inputs and output values are covered by the activation function and last convolution [11].

The authors presented an automatic personality trait recognition model that uses Convolutional Neural Network that trained the information using pre-trained word embedding that is helpful to learning the simplest features from the data without any external dependencies. In CNN architecture, a convolutional layer, followed by number of fully connected layers that is used for feature extraction task [12][18].

In the CNN model has, one embedding layer with input dimension is six. The output of embedding layer is given to

the convolution layer that having sixty four kernels. The output of the convolution layer is given to the maxpooling layer for dimensionality reduction and finally the result is given to the dense layer so we get the classification result. Train our dataset by using the architecture that effectively predicts the behavior of the person.

## V. PROPOSED HYBRID MODEL

Hybrid model is the combination of deep learning models includes convolutional neural network (CNN) and recurrent neural networks (RNNs) that is LSTM or GRU. This model applied to input text, which makes comparatively better result in personality prediction. It is taking benefits of both methods as extracting local features by CNN and long-time dependencies learned via RNN. Because hybrid model is efficient for personality detection by long sequences of input text.

### CNN-RNN FUSION NETWORK

To overcome the problems inside the conventional model and get the better prediction result, we use a hybrid CNN-RNN network. In this network, the purpose of CNN is to learn the local features from input text and RNN takes each words in the input in a sequential order and it is used to learn the long-term dependencies [13]. The network contains following layers: Embedding layer, convolution layer, pooling layer, RNN layer and dense output layer. In RNN layer, using either LSTM or GRU for capturing the long term dependencies in text. The proposed hybrid architecture shown in the following figure 5.1:

#### i. Word Embedding Layer

Word embedding is usually the first hidden layer within the network. The cause of the layer is to transform our input text into vector. Because a text data cannot process with neural network that can be converted into corresponding vectors [14]. The GloVe (Global Vectors for Word Representation) model is used to provide a suite of pre-trained word embedding's for corresponding input text.

#### ii. Convolution and Pooling

A convolutional neural network (CNN) is a type of hierarchical feed forward neural network model, having convolution and pooling operations. In convolutional neural network first we ought to fix the dimensions of the kernel, a kernel may be a multidimensional array that is applied to a window size  $n \times h$ . where  $n$  is that the word vector length of the embedding layer, and  $h$  is that the filter size. The convolutional layers use Rectified Linear Units (ReLU) as activation function [15]. The pooling layer can also be applied over the feature generated from convolution operation that reduce the dimensions of the features.

#### iii. Recurrent Neural Network

The features generated from convolution and pooling operation can be given as the input to the RNN layer, which

has advanced features like n-grams. Therefore, the recurrent neural network (RNN) can process sequential input and learn the long-term dependencies. LSTM or GRU can be applied in this layer based on which model gets better result. When compared with LSTM, the GRU reduces the number of gated units, which decreases the processing time while retaining the accuracy.

#### iv. Fully Connected Network with Softmax Output

The features generated from RNN are passed to a fully connected layer, which is the dense output layer. Any of the classification methods can be used in this output layer. We use softmax classifier in the dense output layer. That is, Softmax assigns probabilities to each class in a multi-class problem. Softmax is implemented in a neural network layer just before the output layer. The Softmax layer must have the same number of nodes as the output layer [16].

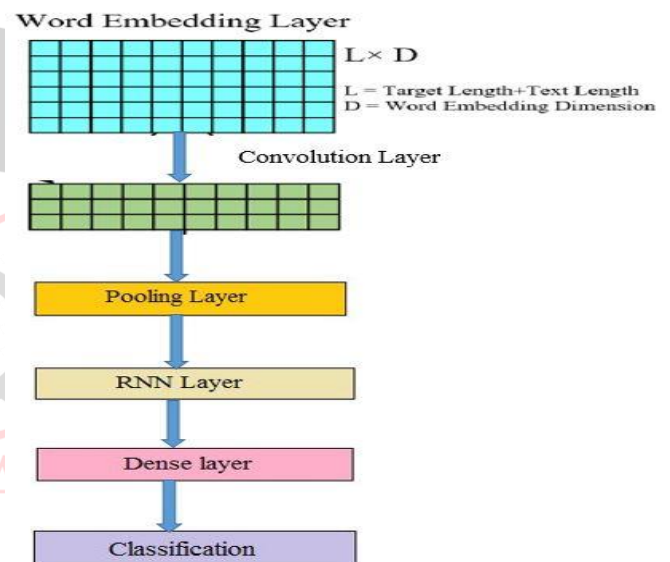


Fig. 5.1 Architecture of Hybrid CNN-RNN network

First, the local features of the data has extracted by the convolution layers with filter window sizes is 128. Then the output of convolution layer is given to the input of GRU layer, it contains 75 nodes and perform another feature extraction. In pooling layer, dimensionality reduction can takes place. The output of maxpooling layer is passed to the dense output layer. In this layer, a classifier can work and predict the result. The prediction of the personality based on big five personality traits is done [17].

Using the features of CNN and LSTM/GRU, it will provide good classification result and accuracy. CNN takes variable length input and process effectively. Another key feature is that deep convolutional networks are flexible and work well with text data. Unlike LSTM, GRU use less training parameters and use less memory that execute and train faster than LSTM. Compare with above prediction models such as, CNN, LSTM and GRU this hybrid version of

CNN-RNN i.e., CNN-LSTM and CNN-GRU network give the better prediction result and training accuracy.

## VI. RESULT AND DISCUSSIONS

The essay dataset contains 2460 independent essays. 60% data is used for training and 40% used for validation [16]. We trained the dataset by using various models such as hybrid CNN-RNN (ie, CNN-LSTM, CNN-GRU), LSTM, GRU and CNN with 20 epochs. Table 6.1 summarizes the result of performance accuracy of these models, which shows, training and testing accuracy and number of trained parameters of the above models.

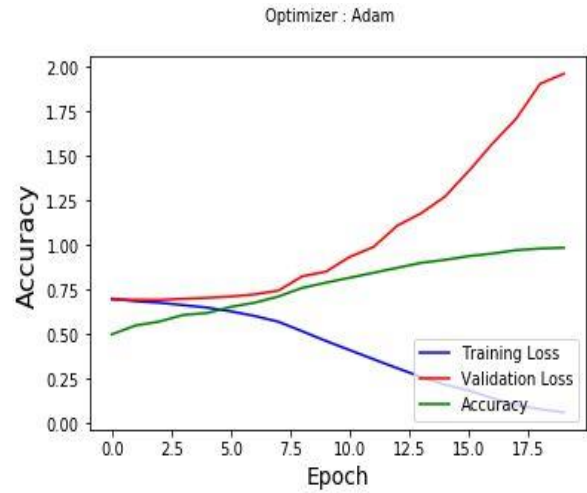
**Table 6.1: Performance summary of different architectures over 20 epochs**

Model	Training accuracy	Testing accuracy	#Params
HYBRID CNN-GRU	99.98	97.1	86658
HYBRID CNN-LSTM	99.24	96.5	61200
CNN	93.37	83.98	57998
GRU	84.1	81.81	43535
LSTM	78.95	74.52	23429

The above table shows the performance summary of different methods such as, Hybrid models and CNN, LSTM and GRU. Accuracy result after 20 epochs of training and validation in Hybrid models are higher than the traditional model. We got 99.98 accuracy for hybrid CNN- GRU model is highest the least accuracy is 78.95 is given by LSTM. In addition, shows the number of parameter developed during each model built. This parameter value control the learning process.

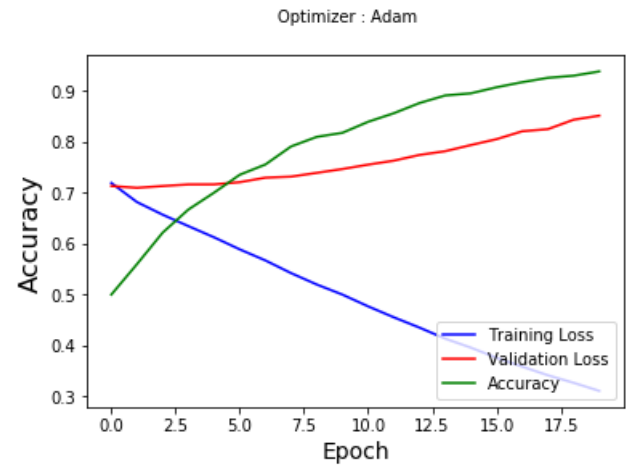
The performance of existing methods and the proposed architecture.

### CNN-GRU network

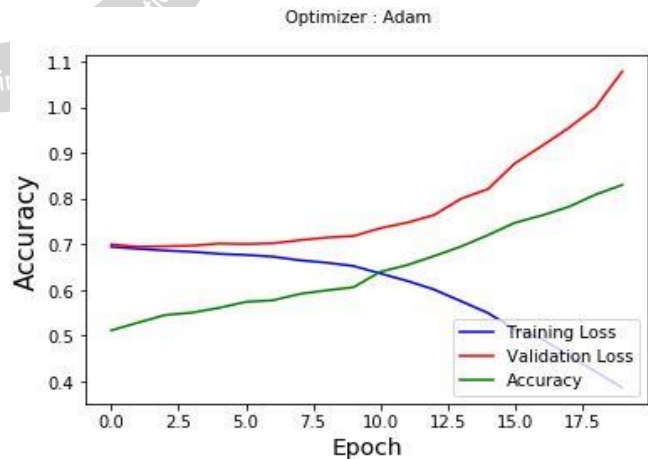


**Fig.6.2 Accuracy plot for hybrid**

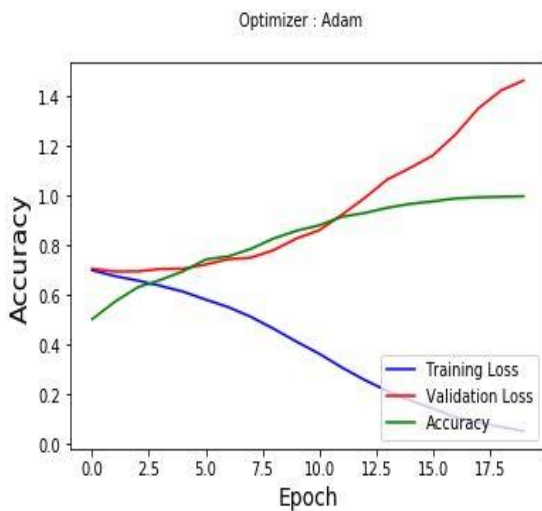
### CNN-LSTM network



**Fig.6.3 Accuracy plot for CNN**



**Fig.6.4 Accuracy plot for GRU**



**Fig.6.1 Accuracy plot for hybrid**

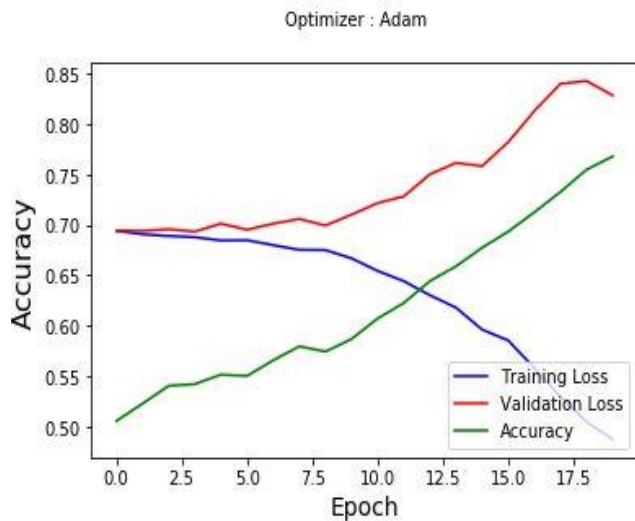


Fig.6.5 Accuracy plot for LSTM

The plot for accuracy, training and validation loss are shown in figures from 6.1 to 6.5. Here, we use Adam optimizer for plotting the graphs this method is quite computationally efficient and easy to implement. And Loss is the sum of errors made for each data in the training and validation. The model is overfitting right from epoch 10 so, the validation loss is increasing while the training loss is decreasing. We got 78.75% accuracy for LSTM, 84.1% accuracy for GRU, 93.37% accuracy for CNN and 99.24% accuracy for CNN-LSTM. Finally, we got 99.98% accuracy for hybrid CNN-GRU network. Compared with the existing models and experimental result, we found that the hybrid CNN-RNN model gave the best accuracy and prediction. Unlike LSTM, GRU reduces the number of gated units and makes faster training because Hybrid CNN-GRU model provided a better personality prediction.

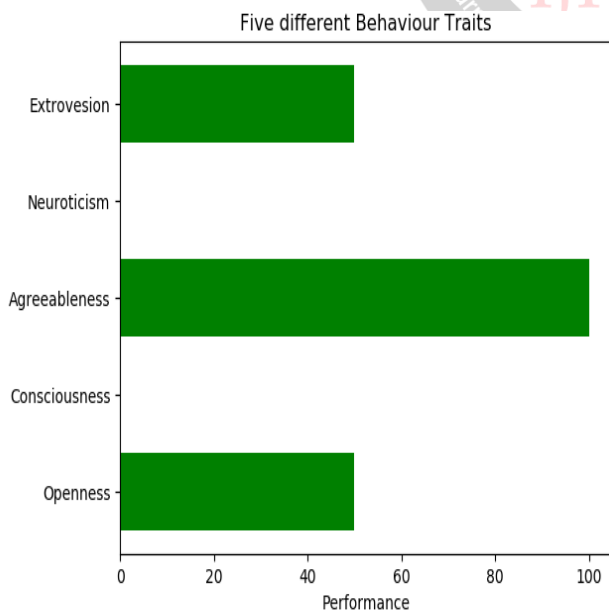


Fig.6.6 Bar chart for user input

Figure 6.6 shows the bar chart for the input given by user. For each of the inputs, taking the unique tokens into consideration, the vector values are generated and the mean

value that goes to the corresponding behavioral trait is taken. The percentage is displayed with the help of a horizontal bar chart. “I am shiny and fast” is the user input that can be process by various models like CNN, LSTM, GRU and hybrid CNN-RNN networks. A person may have different behavioral accents. Based on the user input, it may contain mixed behaviors. The percentage of the obtained behaviors whose value is more than 75 is the expected to be the behavior of that person. Here 50 % accuracy is obtained for extrovesion and openness and 100 % accuracy for agreeableness. So, the behavior of person is agreeableness.

## VII. CONCLUSION

Personality detection is the most important and popular field of natural language processing. The personality detection extract the behavior characteristics of persons that can be used for several applications such as sentiment analysis, HR management, Fraud detection, etc. This paper presents state-of-art review of the emerging fields of personality detection from text using deep learning algorithms and discusses some traditional approaches. Then find which model is better for personality detection. We got A hybrid version of the CNN-RNN network are more suitable model for personality detection especially for long sequences of data when compared with traditional approaches such as LSTM, GRU and CNN. The hybrid model makes faster training and provided better prediction accuracy than the existing techniques

## REFERENCES

- [1] Basant Agarwal , Personality Detection from Text: A Review, International Journal of Computer System, 30(1), Volume 01– Issue 01, September, 2014.
- [2] Hochreiter, S., & Schmidhuber, J. Long short term memory. Neural computation,9(8), 17351780, 1997.
- [3] Mairesse, F. and Walker, M. A. and Mehl, M. R., and Moore, R, K. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. In Journal of Artificial intelligence Research, 30(1), pp: 457–500, 2007.
- [4]Kowsari, K.; Jafari Meimandi, K.; Heidarysafa, M.; Mendu, S.; Barnes, L.; Brown, D. Text Classification Algorithms: A Survey. , vol.10, pp.150, 2019.
- [5] Pennebaker, J. W., King, L. A. Linguistic styles: Language use as an individual difference. In Journal of Personality and Social Psychology, 77, 1999. .
- [6] Celli, F., Pianesi, F., Stillwell, D. S., and Kosinski, M.. Workshop on Computational Personality Recognition (Shared Task). The Seventh International AAAI Conference on Weblogs and Social Media. Boston, MA, USA, 2013.

- [7] PhongLe and Willem Zuidema, Compositional Distributional Semantics with Long Short Term Memory, 2015.
- [8] Navonil Majumder, Soujanya Poria, “Deep Learning Based Document modeling for Personality Detection from Text”, Conference on Affective Computing and Sentiment Analysis, 2017.
- [9] Qing Yu, Hui Zhao, Zuohua Wang, “Attention-based bidirectional gated recurrent unit neural network for sentiment analysis”, 2<sup>nd</sup> International Conference on Artificial Intelligence and Pattern Recognition. ACM ISBN-978-1-4503-7229, pp.116-119, August 2019.
- [10] Dev R& Salemt, F. M. Gate-variants of Gated Recurrent Unit (GRU) neural networks. IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), 2017.
- [11] Maite Giménez, Roberto Paredes, and Paolo Rosso, Personality Recognition Using Convolutional Neural Networks, Springer Nature Switzerland, CICLing 2017, LNCS 10762, pp. 313–323, 2018.
- [12] Y. LeCun, Y. Bengio et al., “Convolutional networks for images, speech, and time series,” The handbook of brain theory and neural networks, vol. 3361, no. 10, 1995.
- [13] Wenfa li, Yilong xu, and Gongming wang, “Stance Detection of Microblog Text based on Two-channel CNN-GRU Fusion Network”, IEEE Access, vol.10, 2017.
- [14] Carducci, G.; Rizzo, G.; Monti, D.; Palumbo, E.; Morisio, M. TwitPersonality: Computing Personality Traits from Tweets Using Word Embeddings and Supervised Learning. Information, vol.9, pp.127, 2018.
- [15] Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 807– 814, 2010.
- [16] Cho, K., van Merriënboer, B., Bahdanau, D., & Bengio, Y. On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259, 2014.
- [17] Trapit Bansal, David Belanger, Andrew McCallum, "Ask the GRU : Multi task learning for deep text recommendations", 10th ACM Conference on Recommender Systems, pp.107-114, September 2016.
- [18] Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio."On the properties of neural machine translation: Encoder-decoder approaches", arXiv preprint arXiv:1409.1259, 2014.