# Personality Prediction Based on tweets of Russo-Ukrainian Conflict


Vivek Sri Krishna Chaitanya Konakalla
*Department of ECE*
*SRM University-AP*
Andhra Pradesh, India
krishnachaitanya_konakalla@srmap.edu.in

Chandra Naga Sai Manikanta Kona
*Department of ECE*
*SRM University-AP*
Andhra Pradesh, India
manikanta_suresh@srmap.edu.in

Tagore Hari Prasad Chintamaneni
*Department of ECE*
*SRM University-AP*
Andhra Pradesh, India
tagorehariprasad_chinatamaneni@srmap.edu.in

Vanaja Boddu
*Department of ECE*
*SRM University-AP*
Andhra Pradesh, India
vanaja_boddu@srmap.edu.in

V Sateeshkrishna Dhuli
*Department of ECE*
*SRM University-AP*
Andhra Pradesh, India
sateeshkrishna.d@srmap.edu.in



***Abstract*—The Russo-Ukrainian conflict has been a highly contentious and protracted geopolitical issue that has garnered significant attention on various social media platforms, particularly Twitter. Online discussion on Twitter has been the main platform for the protracted and polarizing geopolitical conflict between Russia and Ukraine. As the platform generates a lot of user-generated information and allows us to investigate the prospect of using tweets related to the dispute as a method to predict people's personality traits using Twitter posts. To extract and analyse textual information from tweets on the conflict, we used machine learning methods and natural language processing (NLP) approaches in this work. Based on the data that Twitter users shared during the conflict, the main objective of this study is to forecast the personalities of those people. The linguistic and psycholinguistic characteristics were obtained from the preprocessed data and for understanding the personalities we applied Big five factor model (BFFM) on the dataset. With the help of these characteristics and features, the Big Five scores and personality traits are predicted. The machine learning and deep learning algorithms such as Support vector machine (SVM), MLP (Multilayer Perceptron), and RCNN (Region based convolutional neural network) are used to achieve personalities.**



***Index Terms*—Russo-Ukrainian conflict, algorithms, NLP, BFFM, linguistic characteristics, MLP, RCNN.**


## I. INTRODUCTION

In recent years, information has grown exponentially, notably in the form of textual data kinds, along with the rapid growth of social media. Globally, there were 4.8 billion active social media users as of January 2023. Social media is frequently used by people to voice their thoughts on topics including politics, psychology, money, interpersonal relationships, and the environment. They also use it to discuss their daily lives and the welfare of their families. Sometimes people will use these expressions to describe the behavior and characteristics of others. In fact, prior study [1] demonstrates a strong association between social media user personalities and their online behavior. The study of personality using tweets from Twitter has gained popularity among NLP researchers.

Researchers now have access to a plethora of data on people's online behavior thanks to the growing popularity of social media platforms, which may be used to spot trends and patterns in personality traits. Studies reveal that social media profiles can properly predict extraversion and neuroticism among other personality traits. The personality can be described as a style of influence or utilized to distinguish individual persons. Marketers that wish to target clientele groups or psychologists who want to comprehend how personality traits are portrayed online may find this information beneficial. But it's crucial to keep in mind that a person's social media sites just give a fleeting insight into their personality and shouldn't be depended relied as their primary source of information.

According to article [2], numerous psychological analyses were undertaken to try to predict personality. The language people use in tweets may be used to infer a person's level of extraversion, openness, agreeableness, conscientiousness, and emotional stability. The Big Five Factor Model (BFFM), also known as the OCEAN model, which is often used in research, can be used to quantify these characteristics. Based on the findings, numerous academics created a variety of models that enable identifying the traits that determine an individual's personality.

By these models, connections between personality and psychological diseases, job satisfaction and performance [3], and even interpersonal interactions can be found. Because they contain a wealth of information and have millions of users, social networks are a fantastic place to start when researching the personalities of groups.

In this study, we considered a dataset containing the tweets


Identify applicable funding agency here. If none, delete this.

of the Russo-Ukrainian conflict. The data is preprocessed to extract the linguistic characteristics and psycholinguistic characteristics that are useful for accurate predictions, where these obtained characteristics are used for predicting OCEAN scores. The user personalities [4] can be found by the BFFM scores and by the machine learning and deep learning algorithms.

## II. BACKGROUND AND RELATED WORK

In this section, the study of personality trait prediction is covered. This paragraph is broken up into several categories, such as literature on psycholinguistic tools, computational personality prediction traits developed by researchers, and techniques for data collecting, data preprocessing, and classification used in related study fields.

To computationally forecast personality characteristics, researchers categorized the traits and used methods [5] based on machine learning, such as supervised and unsupervised learning models and classification algorithms.

In order to identify the most potent combinations for predicting personality characteristics, we thoroughly investigated several machine learning [17] and deep learning [18] techniques. Our major goal was to develop a reliable model that could predict people's characteristics during the Russo-Ukrainian Conflict with accuracy.

We created a complete big five component model [19], which includes the essential personality qualities of openness, conscientiousness, extraversion, agreeableness, and neuroticism, to strengthen our analysis. We may better understand and anticipate human behavior by using these characteristics, which also provide us a stronger platform for delving into the personalities of those involved in the dispute.

By assessing data from a variety of independent sources, we aimed to clearly connect users' personality traits with their country's responses [20] to the Russo-Ukrainian Conflict. In order to better understand the dynamics at work, our research aimed to illuminate how people interact with the dispute and how they are impacted by it.

According to the article [6], psychologists have long utilized personality analysis to predict outcomes in life.The use of automatic personality feature detection from written communications in a range of domains has lately attracted a lot of interest within the statistical linguistics and the domain of natural language processing. In this work, we follow the evolution of automated personality characterization evaluation on entirely psychological methods to psycholinguistics to more current systems that just use natural language processing, making use of big datasets that were automatically obtained from social media.

The similarity of user-generated texts and the similarity of their personality traits are connected with each other, according to [7] they exhibited a phenomenon. They contend that even without personality databases and supporting networks, networks can nevertheless be constructed based on how closely their produced textual contents resemble interactions between individuals in daily life.

In the article [18], we make an effort to further our analysis of the relationship between personality and the structure of online writings by asking how we may extract this structural attribute from writings with an online source. And in [17], we go through our findings, namely if a personality may be accurately reflected in a structure created from posts on online social networks. And how can you adequately define the organization of statements?

With the use of personality prediction systems and recommendation models, it is now feasible to comprehend human sentiments and interests more fully, as described in [10]. ML algorithms expedite the examination of the survey and decrease the amount of time required to predict the personality feature. In the article [11], they describe how they used a neural network to create an automated model that could perform without a set of questionnaires and instead used photographs as input. Essential response characteristics are then retrieved using NLP after cleaning, and the proper classification methods are then used. The incorporation of relational data into the classification algorithms will be helpful, according to [12], in order to improve the optimization of the findings. Because the dataset affects how each classification algorithm functions.

Using a natural language toolkit (NLTK), from which we can utilize the natural language processing python package, Sentimental Analysis, which may be beneficial to determine the emotional ratings from the phrases of the different individuals, can be developed, as explained in [13]. The authors of [14] have concentrated on a strategy where author essays are utilized as input, ran a typical natural language processing procedure, and then compared the retrieved feature words with conventional personality characteristic words and scored proportionately.

When comparing different machine learning algorithms, Support Vector Machine (SVM) gives more accuracy and computes more quickly; on the other hand, decision tree is used for sequential categorization. This was discussed in [15] where they had reviewed several machine learning algorithms as classifiers for text classification. Therefore, we used two Deep Learning approaches, RCNN and Neural Network, as well as three Machine Learning approaches in this article, including Naive Bayes, Random Forest, and Support Vector Machine.

In addition to RCNN, a deep neural network model called AttRCNN that also incorporates batch normalization and an attention mechanism is presented in [16]. The model itself will carry out the whole vectorization for the single text post.

## III. OVERVIEW OF ALGORITHMS

In this section, several deep learning and machine learning personality prediction algorithms were thoroughly evaluated. We successfully analysed and interpreted a variety of personality traits through sensible application and selection. The objective of these algorithms was to find trends, connections, and insightful information that helps with accurate personality prediction.

1. Gaussian Naive Bayes:

$$P(B|\mathbf{A}) = \frac{P(\mathbf{A}|B) \cdot P(B)}{P(\mathbf{A})} \tag{1}$$

Gaussian Naive Bayes in equation(1), is a probabilistic classifier that predicts the probability of a target variable $B$ given the input feature vector $\mathbf{A}$. It calculates the posterior probability using Baye's theorem, where $P(B)$ is the prior probability, $P(\mathbf{A}|B)$ is the likelihood, and $P(\mathbf{A})$ is the evidence.

2. Random Forest Classifier:

$$F(x) = \frac{1}{N}\sum_{i=1}^{N} T(x;\theta_i) \tag{2}$$

An ensemble machine learning technique called Random Forest Classifier in equation(2), uses many decision trees to provide predictions. It aggregates the predictions of individual decision trees by averaging them, where $T(x;\theta_i)$ represents the prediction of the $i$-th decision tree for the input $x$, and $N$ is the total number of trees in the random forest.

3. Support Vector Classifier (SVC):

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right) \tag{3}$$

Support Vector Classifier in equation(3), is a binary classification algorithm that finds the optimal hyperplane to separate the data points into different classes. It makes predictions based on the sign of the decision function, which is a linear combination of support vectors weighted by the coefficients $\alpha_i$, target labels $y_i$, and a kernel function $K(\mathbf{x}_i, \mathbf{x})$ that measures the similarity between input vectors.

4. Multilayer Perceptron (MLP):

$$\begin{aligned} f(\mathbf{x}) = \sigma\Big(&\mathbf{W}^{(L)} \cdot \sigma(\mathbf{W}^{(L-1)} \cdot (\ldots \\ &\sigma(\mathbf{W}^{(1)} \cdot \mathbf{x} + \mathbf{b}^{(1)}) \ldots) + \mathbf{b}^{(L-1)}) + \mathbf{b}^{(L)}\Big) \end{aligned} \tag{4}$$

Multilayer Perceptron in equation(4), is a neural network based on feedforward propagation that consists of multiple layers of interconnected artificial neurons. Each neuron applies an activation function $\sigma$ to a weighted sum of its inputs. The MLP's output is obtained by propagating the input vector $\mathbf{x}$ through the network, where $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ denote the weights and biases of the $l$-th layer, respectively.

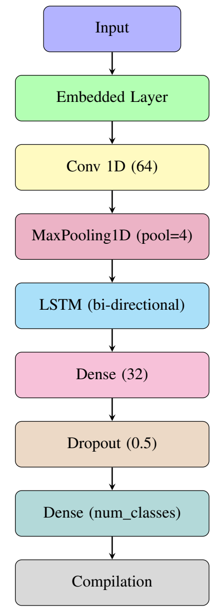


Fig. 1: Flowchart for RCNN Architecture

5. Long Short-Term Memory (LSTM):

$$h_t, c_t = \text{LSTM}(x_t, h_{t-1}, c_{t-1}) \tag{5}$$

Recurrent neural networks in equation(2), that can perceive long-lasting dependencies in sequential input are known as Long Short-Term Memory networks. The internal cell state ($c_t$) and hidden state ($h_t$) are kept up to date depending on the input ($x_t$), the preceding hidden state ($h_{t-1}$), and the ($c_{t-1}$), respectively. LSTM effectively addresses the vanishing gradient problem by incorporating gating mechanisms.

6. Gated Recurrent Unit (GRU):

$$h_t = \text{GRU}(x_t, h_{t-1}) \tag{6}$$

Another kind of recurrent neural network that can recognize sequential patterns is the Gated Recurrent Unit in equation(6). Comparable to LSTM, GRU keeps track of a hidden state $h_t$ that is dependent on the inputs $x_t$ and $h_{t-1}$ before it. It selectively updates and forgets information from earlier time steps by using gating techniques to manage the information flow.

7. Region-based Convolutional Neural Network (RCNN):

The Recurrent Convolutional Neural Network(Fig. 1) is a very potent CNN model that is utilised for processing sequential data, notably in natural language applications like sentiment analysis and named entity recognition. An embedded layer, a Convolution layer (64 filters), a Max pooling layer, and a bi-directional LSTM layer are among the crucial elements of its design that efficiently capture long-term relationships.

While the ReLU activation function is employed across the board, however the activation function of Softmax is applied solely at the output layer. In order to prevent overfitting, a 32-neuron dense layer is included to the model. Next, categorical cross-entropy loss is utilised to create the model, and the RMSprop optimizer serves to determine validation accuracy.

## IV. Methodology

The experimental methodology comprises five stages(Fig. 2), starting with the use of unfiltered Twitter data. Pre-processing, LIWC data processing, and then the formation of a training dataset are the next phases. Then, personality trait scores are assigned to different categories and OCEAN scores are predicted using a Support Vector Classifier (SVC).

### *A. Data Collection*

The dataset used in this study has 29 columns and 47,994 rows, which reflect various characteristics related to each tweet. The first column, titled "Unnamed: 0," looks to be an index or a unique identification for each row. The username of the Twitter account is stored in the "username" column, while the user ID associated with the tweet is stored in the "userid" column.

A few other columns include data like the "text" column contains the tweet's actual content. The language of the tweet is shown in the "language" column, while the "hashtags" column lists any hashtags that were included in the tweet. The dataset includes columns for managing comments, tweeted quotes, and the accompanying user IDs and usernames.

In this study, we concentrate on pertinent characteristics that help in social media personality prediction and classification. We carefully use features from this dataset to carry out the required activities during the analytical and modelling procedures. Additionally, we used a dataset from Kaggle, which contains personality trait scores and matching personality classifications, to train the classifiers.

### *B. Data PreProcessing*

The dataset is typically subjected to the pre-process(Fig. 3). All previously obtained data will be erased if social media statuses contain URLs, symbols, mentions, hashtags, special characters, numbers, punctuation, excessive whitespace, or emoticons. For instance, using must've to become must have is an example of expanding a contraction as part of a phrase. Each sentence is then lowercased to make it sound more natural. Any stop words and clitics will also be deleted in order to prevent ambiguity. The data is pre-processed in this process using the re and string libraries, which provide several linguistic functions to aid in the cleaning up of social media status data.

There will be an additional step during the pre-processing of Twitter data to convert from several languages to only English. The "Lang detect" library is employed in this study to carry out this procedure. We can create a consistent and homogenous dataset using this technique, which will allow us to discover patterns, trends, and insights exclusive to the English language context. By focusing just on English tweets in the dataset, we can more thoroughly investigate the relationship between linguistic patterns and personality characteristics. As a result, we may learn crucial information about how individuals express themselves and exhibit unique characteristics on English-speaking social media platforms.

### *C. Attribute extraction*

This research employs use of the James W. Pennebaker and his team's Linguistic Inquiry and Word Count (LIWC) software, to examine and quantify the linguistic and psychological aspects of texts. In order to divide words into linguistic and psychological categories, LIWC uses a preset vocabulary. This allows researchers to analyse word usage patterns and draw conclusions about the author's psychological and social characteristics.

The following essential features and applications are provided by LIWC:

Linguistic Categories: The LIWC includes several different linguistic categories, such as function words, content words, social words, emotional words, and cognitive words. The numerous facets of language use may be captured by these categories, which also make it easier to analyse word distribution and usage trends.

Psychological Categories: The LIWC also includes psychological categories that represent feelings, social dynamics, thought processes, and personal concerns. These categories' words can be analysed to get knowledge of psychological situations and states.

Text Analysis: LIWC offers a variety of text data analysis techniques, including word category percentage calculations, emotional tone evaluations, time-series analysis of language use, text comparisons, and the identification of linguistic markers linked to psychological characteristics or behaviours.

Psychological Insights: LIWC is widely used in the social sciences, marketing, and communication research. Examining linguistic patterns, social dynamics, cognitive processes, emotional expressiveness, and personality traits in various circumstances is made easier by this. An in-depth understanding of the psychological and social components of text data may be gained by LIWC analysis.

The study extracts linguistic characteristics from the Twitter dataset using LIWC, such as word category distribution, emotional tone, cognitive processes, and social interactions. This technique clarifies patterns and relationships in social media data, which helps us better understand how language usage and personality traits are related. The study's main goals are

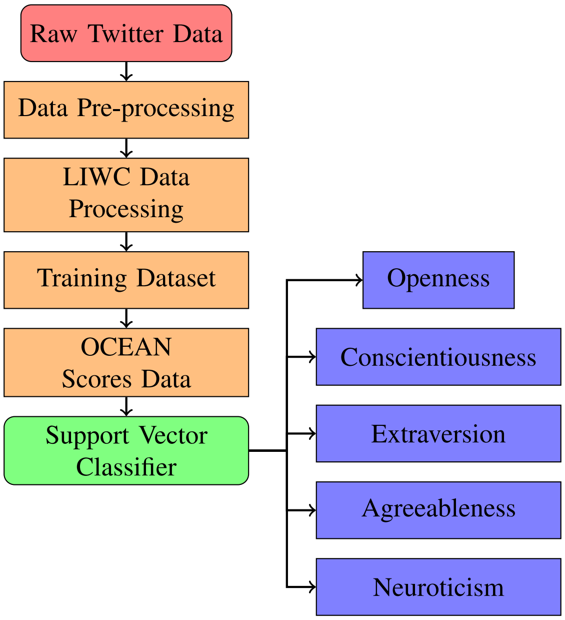


Fig. 2: Flowchart for Methodology

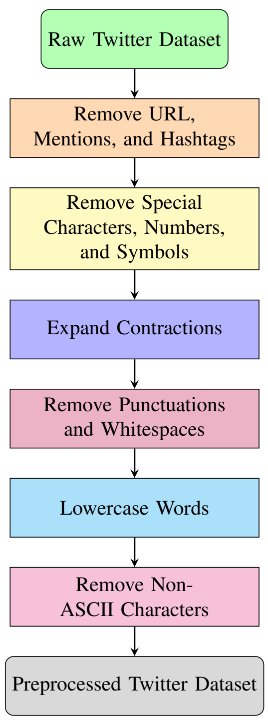


Fig. 3: Flowchart for Twitter Data Pre-processing

furthered by its important contributions to our understanding of language and personality dynamics.

### D. Classification Techniques

A variety of cutting-edge machine learning and deep learning classifiers, including the Gaussian Naive Bayes, Random Forest Classifier, Support Vector Classifier (SVC), MLP (Multilayer Perceptron) classifiers, Long Short-Term Memory (LSTM), GRU (Gated Recurrent Unit), and RCNN (Region-based Convolutional Neural Network), were used in our research. These classifiers were chosen such that measures like precision, recall, F-measure, and accuracy could be used to evaluate how effective they were.

## V. EXPERIMENTAL METHODS

In this section, we evaluate the effectiveness of machine learning models using a range of metrics, including accuracy, precision, recall, and F-measure. An extensive review of how every component of the study was carried out as a single investigation is provided.

The performance of several machine learning methods is shown in the accompanying TABLE I, which also shows the level of accuracy that each model can attain. The Multi-layer Perceptron algorithm had the best accuracy score of 88.11% out of the algorithms tested, demonstrating its efficacy in predicting personality. Contrarily, with an accuracy of 77.06%, the Gaussian Naive Bayes method performed the worst. Personality traits are predicted by using these models, which were trained using OCEAN scores generated from linguistic features.

| Machine Learning algorithms | Metrics | | | |
|---|---|---|---|---|
| | Accuracy (%) | Precision | Recall | F1-score |
| Gaussian Naive Bayes | 77.06 | 0.8673 | 0.771 | 0.7969 |
| Random Forest Classifier | 80.24 | 0.7411 | 0.8024 | 0.7674 |
| Support Vector Classifier (SVC) | 85.99 | 0.7977 | 0.86 | 0.8246 |
| MLP (Multilayer Perceptron) | 88.11 | 0.8454 | 0.8811 | 0.8512 |

TABLE I: Machine Learning algorithms Performance

The TABLE II provides a thorough and extensive examination of several deep learning algorithms used for personality prediction. It includes a variety of assessment criteria, such as accuracy, loss, validation accuracy, and validation loss, and offers a thorough analysis of the algorithms' performance. The Region-based Convolutional Neural Network (RCNN), one of the assessed models, comes out as the most effective one, obtaining an astounding accuracy of 93.96%. Additionally, the LSTM and GRU algorithms, which have accuracy values of 91.43% and 91.67%, respectively, perform well. Our comprehension of the algorithm's general efficacy and robustness is further improved by the table's inclusion of loss and validation measures.

| Deep Learning algorithms | Metrics | | | |
|---|---|---|---|---|
| | Accuracy (%) | Loss | Validation Accuracy | Validation Loss |
| Long Short-Term Memory (LSTM) | 91.43 | 0.2870 | 0.9200 | 0.2881 |
| Gated Recurrent Unit (GRU) | 91.67 | 0.2846 | 0.9149 | 0.2866 |
| Region-based Convolutional Neural Network (RCNN) | 93.96 | 0.1948 | 0.9495 | 0.1783 |

TABLE II: Deep Learning algorithms Performance

## VI. COMPARATIVE ANALYSIS

The effectiveness of the LSTM, GRU, and RCNN models was assessed based on the distinct cases of Loss, Accuracy, Validation Loss, and Validation Accuracy.

In terms of loss(Fig. 4), the LSTM model started with a value of 0.6698 and gradually decreased to 0.2870, while the GRU model began at 0.6138 and reached a lower value of 0.2846. On the other hand, the RCNN model demonstrated impressive performance from the beginning, starting with a higher loss of 0.6377 but swiftly reducing it to 0.1948.

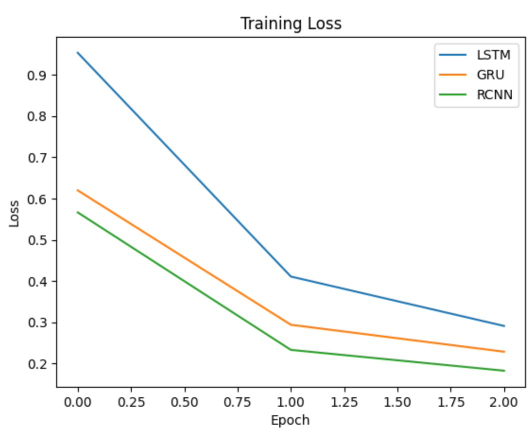


Fig. 4: Training Loss

Moving on to accuracy(Fig. 5), the LSTM model showed consistent improvement, starting at 0.7957 and reaching an accuracy of 0.9143. Similarly, the GRU model exhibited steady progress, starting at 0.8021 and concluding with an accuracy of 0.9167. The RCNN model stood out with the highest accuracy, starting at 0.7673 and achieving an impressive accuracy of 0.9396.

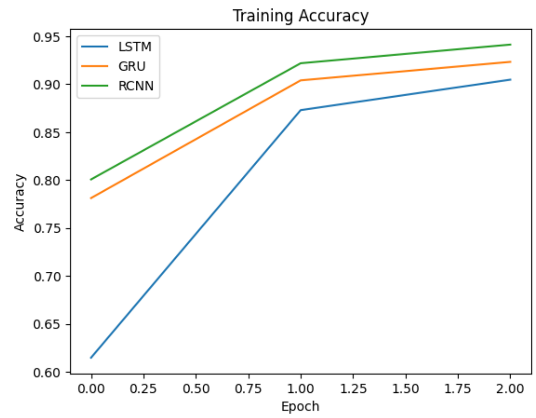


Fig. 5: Training Accuracy

Regarding validation loss(Fig. 6), the LSTM model started at 0.4542 and decreased to 0.2881, while the GRU model began at 0.4274 and reached a validation loss of 0.2866. The RCNN model displayed significant improvement, starting with a higher validation loss of 0.2520 and reducing it to 0.1783.

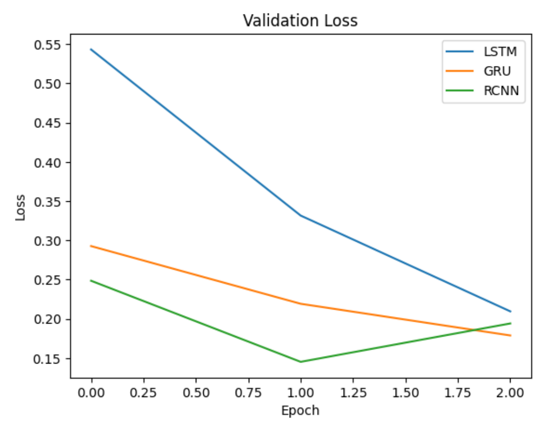


Fig. 6: Validation Loss

Finally, in terms of validation accuracy(Fig. 7), the LSTM model improved from 0.8644 to 0.9200, the GRU model showed consistent performance with values ranging from 0.8750 to 0.9150, and the RCNN model achieved remarkable validation accuracy, starting at 0.9109 and reaching an impressive value of 0.9495.

In the task of predicting personality traits using the OCEAN Scores, the effectiveness of several machine learning algorithms and deep learning models was assessed. Loss and accuracy(Fig. 8) indicators were used as the basis for assessment.

The Support Vector Classifier (SVC), one of the conventional machine learning techniques, has the maximum accuracy of 85.99% with a corresponding loss of 0.8246. The MLP

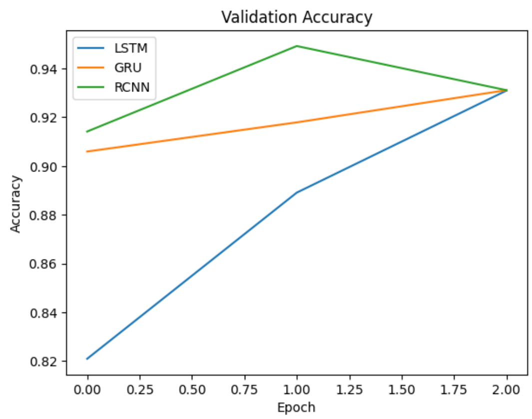


Fig. 7: Validation Accuracy

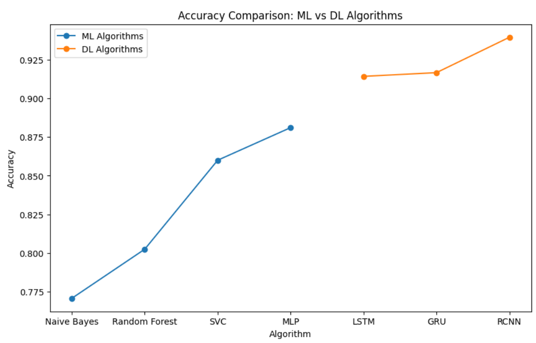


Fig. 8: Comparison of ML and DL Algorithms

Classifier also showed noteworthy accuracy, achieving 88.11% with a loss of 0.8512. The Random Forest Classifier has an accuracy of 80.24% and a loss of 0.7674. In contrast, the Naive Bayes Classifier had a lower accuracy of 77.06% and a loss of 0.7969.

The Long Short-Term Memory (LSTM) model performed well among deep learning models, with an accuracy of 91.43% and a loss of 0.2870. The accuracy was 91.67% with a loss of 0.2846 for the Gated Recurrent Unit (GRU) model, which was marginally more successful. Notably, the Region-based Convolutional Neural Network (RCNN) model surpassed all other methods with an excellent accuracy of 93.96% and a noticeably lower loss of 0.1948.

When comparing the outcomes, it was found that the deep learning models outperformed the conventional machine learning methods in terms of accuracy and loss. This demonstrates how well deep learning approaches work for predicting personality traits based on language characteristics. The greater accuracy levels attained by the deep learning models, the RCNN model, demonstrate their capacity to identify complex connections and patterns in the data.

## VII. CONCLUSION

Finally, the personalities of Twitter users who engaged in the Russian-Ukrainian war were anticipated by our work. The outcomes showed that the RCNN model had the best level of accuracy, coming in at 94%. The Gaussian Naive Bayes algorithm, on the other hand, performed comparatively less accurately, achieving an accuracy of 77%. All results of these models were evaluated with Accuracy, Precision, Recall, and F1-score.

The effect of linguist characteristics derived from the LIWC software showing the efficiency of these qualities was found to be restricted in some situations where the information about the tweet was unclear, even though they offered useful insights into several characteristics. This implies that adding more features or include more contextual data could increase the prediction precision in certain situations.

## VIII. ACKNOWLEDGEMENT

I would like to thank my mentor Dr. Sateesh Krishna Dhuli for his valuable guidance and support for my research work. My heartfelt appreciation to the seniors for investing their time and effort in helping me. I am truly grateful for their mentorship and the opportunity to learn them.

## IX. REFERENCES

[1] Ahmed Al Marouf, Md. Kamrul Hasan, and Hasan Mahmud, "Comparative Analysis of Feature Selection Algorithms for Computational Personality Prediction From Social Media," IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS, VOL. 7, NO. 3, JUNE 2020.

[2] Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner, "2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing."

[3] Timothy A Judge, Joyce E Bono, and Edwin Locke, "Personality and job satisfaction: the mediating role of job characteristics," Volume 85, Number 2, Year 2000.

[4] William Villegas-Ch, Daniel Mauricio Erazo, Ivan Ortiz-Garces, Walter Gaibor-Naranjo, and Xavier Palacios-Pacheco, "Artificial Intelligence Model for the Identification of the Personality of Twitter Users through the Analysis of Their Behavior in the Social Network," Volume 11, Number 22, Year 2022.

[5] International Personality Item Pool. Accessed: Jan. 7, 2020. [Online]. Available: https://ipip.ori.org/

[6] Alireza Souri, Shafgheh Hosseinpourand Amir Masoud Rahmani, Personality classification based on profiles of social networks' users and the five-factor model of personality Souri et al. Hum. Cent. Comput. Inf. Sci. (2018) 8:24 https://doi.org/10.1186/s13673-018-0147-4

[7] Liyang Gao, Weiguo Wu, Relevance assignation feature selection method based on mutual information for machine learning Knowledge-Based Systems. Volume 209, 17 December 2020, 106439

[8] X. Sun, B. Liu, J. Cao, J. Luo and X. Shen, "Who Am I? Personality Detection Based on Deep Learning for Texts," 2018 IEEE International Conference on Communications (ICC), Kansas City, MO, USA, 2018, pp. 1-6, doi: 10.1109/ICC.2018.8422105

[9] Gita Safitri, Erwin Budi Setiawan Optimization Prediction of Big Five Personality in Twitter Users, Vol 6 No 1 (2022): February 2022 Section Information Systems Engineering Articles

[10] P. D. J. Tejas Bakade, Siddhant Deshpande, Omkar Pethe, Shreyas Kale, Yashwant Madane, "Big Five Personality Prediction Using Machine Learning Algorithms," MSEA, vol. 71, no. 3, pp. 1128 -, Aug. 2022.

[11] T. Pradhan, R. Bhansali, D. Chandnani and A. Pangaonkar, "Analysis of Personality Traits using Natural Language Processing and Deep Learning," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2020, pp. 457-461, doi: 10.1109/ICIRCA48905.2020.9183090.

[12] Kadhim, A.I. Survey on supervised machine learning techniques for automatic text classification. Artif Intell Rev 52, 273-292 (2019). https://doi.org/10.1007/s10462-018-09677-1

[13] Jiawei Yao. (2019). Automated Sentiment Analysis of Text Data with NLTK. Journal of Physics: Conference Series, 1187(5), 052020. doi: 10.1088/1742-6596/1187/5/052020.

[14] K. C. Pramodh and Y. Vijayalata, "Automatic personality recognition of authors using big five-factor model," 2016 IEEE International Conference on Advances in Computer Applications (ICACA), Coimbatore, India, 2016, pp. 32-37, doi: 10.1109/ICACA.2016.7887919.

[15] A. Bhavani and B. Santhosh Kumar, "A Review of State Art of Text Classification Algorithms," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 1484-1490, doi: 10.1109/ICCMC51019.2021.9418262.

[16] Zhuoyao Zhong, Lei Sun, and Qiang Huo. (2019). An anchor-free region proposal network for Faster R-CNN-based text detection approaches. International Journal on Document Analysis and Recognition (IJDAR), 22(3), 315-327. doi: 10.1007/s10032-019-00335-y

[17] Ergu, İzel and Işık, Zerrin and Yankayış, İsmail. Predicting personality with Twitter data and machine learning models.

[18] Christian, Hans and Suhartono, Derwin and Chowanda, Andry and Zamli, Kamal Z. Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging. Volume 8, Number 1, Year 2021.

[19] Golbeck, Jennifer and Robles, Cristina and Edmondson, Michon and Turner, Karen. Predicting personality from Twitter.

[20] Gil de Zúniga, Homero and Diehl, Trevor and Huber, Brigitte and Liu, James. Personality traits and social media use in 20 countries: How personality relates to the frequency of social media use, social media news use, and social media use for social interaction. Volume 20, Number 9, Year 2017.