Journal of Al-Qadisiyah for Computer Science and Mathematics Vol.16.(4) 2024,pp.Comp 231–250

University Of AL-Qadisiyah

Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)

Random Multimodal Convolutional Forward Taylor Network for Personality Prediction using MBTI Data

Vian Sabeeh[a], Ahmed Bahaaulddin A. Alwahhabb[b]

[a,b]Department of Information Management Technologies, Technical College of Management-Baghdad, Middle Technical University, Baghdad, Iraq.
Email: (viantalal@mtu.edu.iq), Email:(ahmedbahaaulddin@mtu.edu.iq)

ARTICLE INFO

Article history:
Received: 19 /10/2024
Rrevised form: 22 /11/2024
Accepted : 11 /12/2024
Available online: 30 /12/2024

Keywords:

Personality Prediction;
Random Multimodal Deep Learning;
Convolutional Neural Network Taylor Series;
Deep Learning;

ABSTRACT

The growth of social media for self-awareness and self-expression has raised more attention to the Myers-Briggs Type Indicator (MBTI) for analyzing people's personalities. Nevertheless, additional research is required to determine various word embedding and data handling methods to enhance the accuracy of MBTI personality-type predictions. Therefore, a new technique known as Random Multimodal Convolutional Forward Taylor Network (RMConv FT-Net) is devised to predict personality. Initially, text data from the database is considered input, and then tokenization is performed to split the text into tokens, which is done by Bidirectional Encoder Representations from Transformers (BERT). After this, various features, including linguistic features and several other features, like Term Frequency-Inverse Document Frequency (TF-IDF), Word2Vec, and so on, are mined from the text. Following this, the data imbalance problem is addressed by employing oversampling. Finally, personality prediction is performed by utilizing RMConv FT-Net, designed with the incorporation of the Convolutional Neural Network (CNN), Taylor series, and Random Multimodal Deep Learning (RMDL). The experimental outcomes of RMConv FT-Net based on Mean Square Error (MSE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) show that it obtained the values of 0.017, 0.047, and 0.129.

MSC..

https://doi.org/ 10.29304/jqcsm.2024.16.41785

1. Introduction

Techniques for Natural Language Processing (NLP) are becoming increasingly common in organizational and industrial psychology [1]. The investigation and practice of NLP focuses on the practical applications of computer-aided understanding and speech recognition. To develop appropriate tools and techniques for computer systems to understand and manipulate natural languages to perform the necessary tasks, NLP research aims to gain more information about how people perceive and use language. NLP is based on several fields of study, including psychology, artificial intelligence, electrical and electronic engineering, robotics, linguistics, mathematics, computer and information sciences, etc. NLP is used in many research fields, including speech recognition, artificial

∗Corresponding author Vian Sabeeh

Email addresses: viantalal@mtu.edu.iq

Communicated by ‘sub etitor’

intelligence and expert systems, machine translation, natural language text processing and summarization, user interfaces, multilingual and Cross-Language Information Retrieval (CLIR), and so on [2]. When performing content analysis in NLP, a technique is tuned to automatically allocate scale items to the constructs they measure. This process is known as text classification [1].

People's behavior, motivation, feelings, and mental patterns combine to form their personality [3]—theories in personality psychology attempt to provide a reliable and concrete explanation for human personality [4]. Our personalities enormously impact our lives, affecting our preferences, desires, well-being, health, and life decisions. Therefore, there are numerous significant real-world uses for the capacity to identify a person's personality traits [3] automatically.

Traditionally, personality assessments have been conducted using traditional techniques like questionnaires or interviews, which are expensive, time-consuming, and imprecise. However, with the rapid development of online social networks in recent years, there has been an increased interest in research on automatic personality recognition from these platforms. New methods are being developed every day [5]. In any company, personality prediction is crucial to the hiring process because it gives the hiring authority an overview of the candidate's potential value. Understanding the user's age, gender, personal traits, occupation, and political preferences is made easier with the help of personality prediction. Personality prediction is helpful in various fields, including hiring, consulting, business, behavioral analysis, digital marketing, and e-commerce [6]. In particular, the prediction of personality traits from multimodal data has become a popular subject in affective computing [3]. Current motivational research highlights personality's forward-looking, self-constructive elements regarding objectives and future intentions. Although comparatively at how explicit motives relate to one another, explicit motives, such as life tasks, personal projects, and personal strivings, represent what people want or are trying to accomplish in their lives [7]. Social media personality predictions are limited to describing the individual user's personality. Predictability studies can be performed on a particular sample even though this estimation is less accurate than the data gathered in a clinical setting. Social media content can be used for online marketing, hiring procedures, recommendation engines, advertising agencies, criminal justice, intelligence profiling, and other purposes where personality traits can be extracted and predicted [8].

Numerous studies investigated the prediction of an individual's personality traits using standard Machine Learning (ML) algorithms and typical lexical features. However, the performance of these approaches could be much higher to develop a reliable system that can identify personality traits. The limitations of the machine learning model are caused by its dependency on data quantity and quality; in other words, limited or biased data may result in an inaccurate model. Additionally, the choice of features used for training the model can significantly impact its performance. In personality prediction, selecting the right textual features that correlate with personality traits is critical, and poor feature selection can lead to suboptimal results. However, Machine learning models may need help generalizing among various populations or ethnicities. A model trained on one social media platform may need to improve on another due to language use and user behavior differences. Due to this, other researchers have chosen to use advanced methods for ML, like Deep Learning (DL), which can significantly aid in personality prediction by extracting syntactic and semantic information from user posts and implementing novel concepts for modeling sentences and documents [3]. Several research works have been conducted to investigate the best feature space and ML algorithms for personality recognition. Still, modern methods have yet to produce satisfactory results. Furthermore, it has been shown in recent years that distributed representation and DL-based Neural Networks (NN) are effective in modeling sentences and documents and have produced impressive results in NLP applications, like sentiment analysis and opinion mining using text. Additionally, these NLP applications share similarities with our personality recognition task in that they both entail extracting user attributes from texts; the most similar challenge between them is the feature representation tasks [6][9]. DL-based models can help with various tasks such as computer vision, NLP, speech recognition, and handwriting generation. Various DL frameworks, namely Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Gated Recurrent Unit (GRU), and Long Short-Term Memory (LSTM), have been used to classify personalities. Better ways to capture semantic information are impossible for these individual DL models. DL models' capacity for self-learning allows them to outperform ML classifiers that rely on traditional features regarding accuracy and processing speed [10].

Since Personality prediction from text data has gained considerable interest, it has become complicated due to its implications in multiple fields, including psychology, opinion mining, and sentiment analysis. Various researchers have concentrated on features engineering, namely psycholinguistic datasets and linguistic styles, which have many connections with personality. Moreover, transfer learning in feature extraction and fine-tuning trained models has substantially impacted NLP recently.

This paper tackles the problem of developing a reasonably accurate personality prediction system that overcomes imbalanced data issues in MBTI data and uses various linguistic features with optimized convolutional forward Taylor to overcome ML limitations. Input text data from the database is tokenized to split the texts into tokens, which BERT does. After this, various features, including linguistic features and several other features, like TF, TF-IDF, Word2Vec, and so on, are mined from the text. Following this, data imbalance handling is carried out by employing oversampling. At last, personality prediction is performed by utilizing RMConv FT-Net, which is formed by incorporating CNN, Taylor series, and RMDL. The contribution of this paper is:

- Propose a new model for personality prediction. The proposed model, RMConv FT-Net, is utilized to achieve this task. It is produced by combining RMDL, CNN, and Taylor series.

The subsequent sections in this paper are arranged as follows: Section 2 shows the merits and demerits of existing models. Section 3 describes the proposed methodology RMConv FT-Net for personality prediction. Section 4 deliberates the results and discussion of RMConv FT-Net, and Section 5 illustrates the conclusion.

2. Related Work

Due to the importance of personality prediction in various domains, researchers conducted enormous studies to develop reliable and accurate models that depend on machine learning and deep learning models. For instance, El-Demerdash et al. [11] established a DL-based fusion model for personality prediction. This method achieved maximum accuracy independent of the language and psycholinguistic features employed, and it worked well even without an external feature set. Nevertheless, this method was not extended to other affective concepts and subjective terms, like sentiment, opinions, emotion, and mood. Ryan, G. et al. [12] developed Logistic Regression (LR) integrated with Word2Vec and Synthetic Minority Oversampling Technique (SMOTE) for personality prediction. This approach was very flexible and reduced data noise, which helped to solve the imbalanced data problem. However, this approach should have considered deep learning models for improving the recognition of personality types. Serrano-Guerrero J. et al. [5] devised a Stacked Ensemble framework for personality trait prediction. This framework solved the classification and regression issues. Moreover, the model's variety was enhanced, decreasing bias and variation. However, this framework should have included transformer-based models to use contextual information better.

While. Rangra, K. et al. [13] introduced Mel Frequency Cepstral Constants (MFCC) Particle Swarm Optimization (PSO) based Convolution Neural Network (CNN) (NPSO) (MFCC-based NPSO_CNN) for emotional speech-based personality prediction. This model accurately described and correctly identified personality behaviors and feelings from the audio. Nevertheless, misclassifying emotions as neutral, depressed, and angry for the new sound samples made the results unacceptable. Personality prediction relying on audio may be sensitive to context and environment because the recordings can be affected by background noise, emotional state, and speaker style, resulting in the problem of distinguishing personality traits from situational conditions. Moreover, the audio-based system can reveal extroversion or agreeableness that is expressed easily using vocal features. In contrast, more profound aspects of the personality, such as conscientiousness or openness to experience, may be harder to infer from audio alone.

Ramezani, M. et al. [14] established a Bidirectional Long Short-Term Memory (BiLSTM)-based classifier for automatic personality prediction. This method provided more training capacity by traversing the input data twice and improving the prediction rate. Nevertheless, the computational complexity of this technique was high, and convergence was slower. However, despite the high performance of the deep learning method, it performed only moderately due to the limited amount of personality-related content in the essays used for training. Kosan, M.A. et al. [8] devised an LSTM-based Neural Network for predicting personality traits. Using an optimally distributed, high-dimensional dataset, this model produced a balanced success rate. The complexity and volume of processed data in this model increase the hardware constraints. Others used the paradigm of model fusion to build a hybrid model, like Bhagat A. et al. [6], who developed the Big Five model to predict the candidate's personality. This method had a minimum loss function and error in the training and validation stages. Nevertheless, this method's attempt to use different data scraping techniques to separate the actual data from the fake data was unsuccessful. Another work by Ramezani M. et al. [15] introduced the Term Frequency Vector-based method for automatic personality prediction. The training process for this technique took less time. However, real-time scenarios were not used to apply this technique.

As a result of previous studies, we summarized the most critical problems faced by earlier researchers in predicting personality as follows:

- An effectual technique named the DL-based fusion method was developed by El-Demerdash et al. [11] for personality prediction. This approach effectively increased performance and generalizability. However, this model did not consider optimum features for precise personality prediction.
- The LR+SMOTE in Ryan et al. [12] failed to incorporate DL algorithms, such as CNNs and RNNs, to predict MBTI personality types more accurately.
- Also, Serrano-Guerrero et al. [5] devised a Stacked ensemble framework for personality trait prediction. This technique offered stability and accuracy to the learning algorithm, minimizing the generalization error of the prediction task. Nevertheless, this method was highly computationally complex and expensive.
- Rangra et al. The MFCC-based NPSO_CNN developed in [13] neglected to utilize text and images for analyzing the performance and deriving the actual short-term personality from the emotional behavior of the speaker.

Previous studies have examined the application of standard lexical features, traditional machine learning algorithms, and deep learning algorithms to predict an individual's personality traits; however, their performance was insufficient in developing a reliable system to identify personality traits.

3. Research Method

This research proposes a novel approach for Personality Prediction using MBTI data. Firstly, input text data is attained from the database, and then BERT implements tokenization to split the texts into tokens [16]. After that, feature extraction is performed by including linguistic features, namely the rate of misspelling, the average number of words per sentence, maximum content similarity, and the percent of numeric and capital words. Other features like TF-IDF, Word2Vec, TF, number of numerical values and punctuation marks, capitalized words, Lin similarity score, length of text, nouns, hashtags, elongated units, and sentiment-based features are mined [17][18]. Next, data imbalance handling is carried out by employing oversampling. Lastly, Personality Prediction is performed in terms of RMConv FT-Net, which is created by the integration of RMDL [19], CNN [20], and the Taylor series [21]. Fig. 1. portrays the schematic view of RMConv FT-Net for Personality Prediction.

3.1. Data acquisition

A dataset that contains r records is considered to predict personality, and the below equation models it,

$$X = \{X_1, X_2, \ldots, X_k, \ldots X_r\} \quad (1)$$

Where X is the whole dataset, r symbolizes an entire quantity of records in the dataset and X_k exemplifies the k^{th} row or record in the dataset.

3.2. BERT tokenization

Bert tokenization [16] is utilized to change the data into tokens, executed by considering X_r as input; BERT comprises a multi-layer bidirectional transformer network that converts the data into discrete words. Tokenization is executed by fine-tuning the personality framework on an unlabeled document to understand the personality. BERT is regarded as a progressive and complex approach in which a personality is analyzed as a token chain, and every token is compared to every other token in the chain to obtain context-specific information and to know its dynamics. The output attained from BERT tokenization is designated as η_w Which is passed on as input for mining the features.

3.3. Feature extraction

Several features, such as linguistic features and many others, are considered for feature extraction, which is done by taking. η_wAs input.

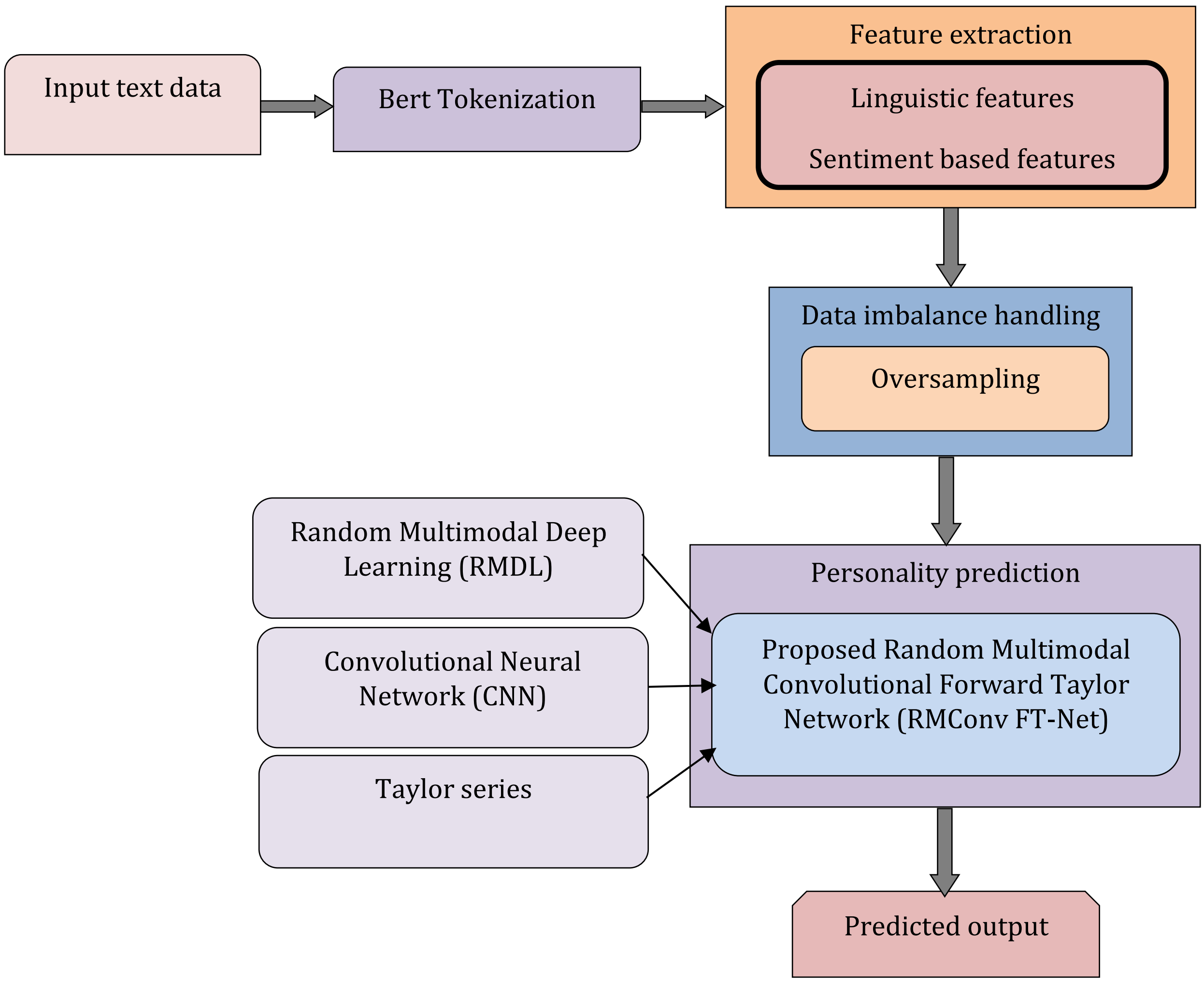


Fig. 1- Schematic view of RMConv FT-Net for Personality Prediction.

3.3.1. Linguistic features

Linguistic features specify the unique qualities of personality, which aids in understanding the context and determining personality. Here, several linguistic features, including misspelling rate, maximum content similarity, average number of words per sentence, and percent of numeric and capital words, are employed, explained below.

1. Average no. of words per sentence

The average number of words [22] in each sentence varies from 15 to 20 words, and this sentence length is readable and offers sufficient information. However, depending on the subject, style of writing, and audience, the size of a sentence varies. Here, T_1 indicates the average number of words in a sentence.

2. Rate of misspelling

The percentage of misspelled words is compared with the entire number of words, which is employed for determining the misspelling rate, and misspelling rate is mentioned as T_2.

3. Percent of numeric

The ratio of the overall number of words in review to the overall numerical employed is calculated to determine the percent of numeric, and it is modeled as [22],

$$T_3 = \frac{W_{num}}{W_h} \qquad (2)$$

Here, the numeric word feature is symbolized as T_3 and its value is ranged between 0 and 1. While W_{num} represents the numerical words in the text and W_h is implied as an entire number of words.

4. Percent of capital words

This feature [4] is determined by dividing the quantity of capital words by the overall number of words, and it is expressed as,

$$T_4 = \frac{W_s}{W_h} \qquad (3)$$

where capital words are exemplified as W_s and the percentage of capital words feature is epitomized as T_4.

5. Maximum content similarity

Similarity among reviews posted by the same user for different products is considered spam, and the cosine similarity is employed for calculating the similarities in content [23], and it is depicted as,

$$T_5 = max_{g_j g_o \in V_t, j<o} \, cos(g_j g_o) \qquad (4)$$

Here, the maximal content similarity is denoted as T_5, Reviews sent by the user are mentioned as V_t, and g_j embodies j^{th} review sent by the user.

3.3.2. Other features

Several other features, namely Word2Vec, text length, elongated units, hashtags, Lin similarity, capitalized words, TF-IDF, nouns, and the number of numerical and punctuation marks, enhance the prediction as follows:

1. TF-IDF

TF-IDF [17] with high value designates that the same word appears more commonly in the review, and this feature is utilized for assessing the importance of a word to a document among corpus. It produces words with relative frequency in the exact document associated with a reverse portion of the word in the entire document group. Moreover, TF-IDF is designated as,

$$T_6 = B_{d_l * b_\omega} \, log\left(\frac{|S|}{B_{d_l * S}}\right) \qquad (5)$$

Equation (5) works to find the value of T6 that calculates the review percentage. b_ω in the document d_l in the whole document by multiplying it with the TF-IDF value that is denoted by $log\left(\frac{|S|}{B_{d_l * S}}\right)$. Here, corpus size is represented as $|S|$, frequency of occurrence of the word d_l in the corpus S and b_ω review is implied as $B_{d_l * S}$ and $B_{d_l * b_\omega}$.

2. Word 2Vec

In NLP, a Word2vec approach [17] is employed for attaining words in vector representations. A text in the large corpus is modeled, and a collection of information about word meanings and context usage is used to predict word representations. Moreover, this approach utilizes the Cosine Similarity equation for calculating the word vector value, and it is represented by

$$Similarity = \cos\phi = \frac{\overline{k}\cdot\overline{v}}{\|\overline{k}\|\|\overline{v}\|} \quad (6)$$

Here, the system value is indicated as k, long vector$k$is characterized as $\|\overline{k}\|$, vector dot product from $k$and$v$ is denoted as $\overline{k}\cdot\overline{v}$, long vector$v$ is symbolized as $\|\overline{v}\|$, The gold standard is mentioned as v, and T_7 exemplified Word2Vec.

3. Number of numerical values

This feature [24] is employed to indicate the total quantity of numerical value in review, and it is written as,

$$T_8 = \sum_{p=1}^{m} L_{\sigma}^{p} \quad (7)$$

where, T_8 specifies numerical value, the overall number of words in p^{th} is signified as L_{σ}^{p},and m indicates the overall number of words in the review.

4. Number of punctuation marks

This feature [17] states the occurrence of punctuations, namely exclamation marks, apostrophes, dots, and so on, in input review. Moreover, it is articulated as,

$$T_9 = \sum_{p=1}^{m} \ell_{Pun}^{p} \quad (8)$$

Here, the number of punctuation marks is denoted as T_9, Count of punctuation in p^{th} review is exemplified as ℓ_{Pun}^{p}.

5. Length of text

The count of words in a review document is termed as the length of the text, and it is mentioned as T_{10}.

6. Lin similarity

According to the information content of reviews, the Lin Similarity Score [25] is calculated in terms of semantic similarity between two reviews, and it is represented as,

$$T_{11} = \frac{2*\gamma\left(E\left(g_j, g_o\right)\right)}{\left(\gamma\left(g_j\right)+\gamma\left(g_o\right)\right)} \quad (9)$$

Here, the Lowest Common Subsumed is indicated as E , information content is epitomized as γ, and Lin similarity score among reviews $g_j$and g_o is signified as T_{11}.

7. Elongated units

The character that appears various times in a review is referred to as an elongated unit [17], and it is expressed as,

$$T_{12} = \sum_{p=1}^{m} N_e^{p} \quad (10)$$

where, T_{12} shows elongated units and the number of elongated units in p^{th} review, which is typified as N_e^{p}.

8. Nouns

An entity named in a review is employed to recognize the essential entities in the review. Moreover, the named entities include ideas, concepts, people, places, and things. Here, a noun is represented as T_{13}.

9. Hashtags

The hashtag feature specifies the number of hashtags in a review; its expression is below.

$$T_{14} = \sum_{u=1}^{H} \ell_{\nu}^{u} \qquad (11)$$

Here, the total count of hashtags is symbolized as ℓ_{ν}^{u}. The Hashtag feature is denoted as T_{14}.

10. Capitalized words

The overall number of capitalized words in a review is termed as capitalized words [17], and it is signified as,

$$T_{15} = \sum_{p=1}^{m} L_{f}^{p} \qquad (12)$$

where, T_{15} showed the capitalized words and the total number of the capital words in the p^{th} review is determined as L_{f}^{p}.

3.3.3. Sentiment based features

A glossary of opinion data attained from the WordNet dataset is provided by sentiment-based features [18]. A polarity score, namely neutral, negative, or positive, is allocated to various groups of synonyms inside words, which are mentioned as synsets. Moreover, the score value is obtained between the range [0,1]. Nevertheless, the sum of scores for every system should be the same, and thus, the opinion is decided. Here, the expression of SentiWordNet is written as,

$$T_{16} = [Y^{c}(\kappa), Y^{c}(\vartheta), Y^{c}(\tau)] = \wp(\chi_{c}) \qquad (13)$$

where, $Y^{c}(\vartheta)$, $Y^{c}(\tau)$ and $Y^{c}(\kappa)$ are implied negative, neutral, and positive scores, while the SentiWordNet function is shown as $\wp$, T_{16} which specified the SentiWordNet feature and χ_{c} presents c^{th}word. Further, the mined features are integrated to attain a feature vector, and it is articulated as,

$$T = \{T_1, T_2, \ldots, T_{16}\} \qquad (14)$$

3.4. Data imbalance

The extracted features T are considered input for addressing the data imbalance issue, which is handled by augmenting the minority class samples using the oversampling technique. This process is essential for minimizing and avoiding overfitting. It is performed as a regulation technique in neural networks to enhance performance in severe classes by producing synthetic new data points. Synthetic Minority Oversampling Technique (SMOTE) has been used in the proposed model due to its ability to handle the class imbalance in the MBTI dataset and many other applications [12]; moreover, to avoid getting biased models that executed poorly on minority classes. Another reason to use SMOTE was to create new text that is similar to the original samples with slight variations, which improved the performance on unseen data, especially for minority classes, which may enhance the generalization of the model and reduce overfitting. As a result, after applying SMOTE to the input features, the new augmented balanced data is generated and mentioned as Γ.

3.5. Personality prediction

A person's personality is a distinctive characteristic that enables individual differentiation. It is characterized by a person's consistent behaviors, which influence their psychological well-being, interactions, relationships, behaviors, and attitudes. Personality prediction has recently received much interest because of the rise of social networking sites offering user-generated text content. These texts from social networks capture users' psychological activity, making them an essential source of data for evaluating the individual characteristics of users. Moreover, RMConv FT-Net is employed for predicting the personalities of the user, where RMConv FT-Net is attained by fusing RMDL [19], CNN [20], and the Taylor series [21]. This RMConv FT-Net includes RMDL, CNN, and RMConv FT-Net layer.

Here, the multiplication of the augmented data Γ with weight J_1 is first executed, and the consequent outcome is normalized. Simultaneously, augmented data ΓIs given to RMDL, and the generated output is multiplied by the normalized outcomes summation fore $\sum\sum J_1\Gamma$ for obtaining γ_1. On the other hand, augmented data Γ is passed to CNN, and the attained outcome is multiplied by the weight J_2 which multiplied with γ_1 for attaining the output γ_2. Finally, the output γ_1 and γ_1 are merged using the Taylor series [24] to attain the outcome γ_3 . Fig. 2. portrays the structural representation of RMConv FT-Net for personality prediction.

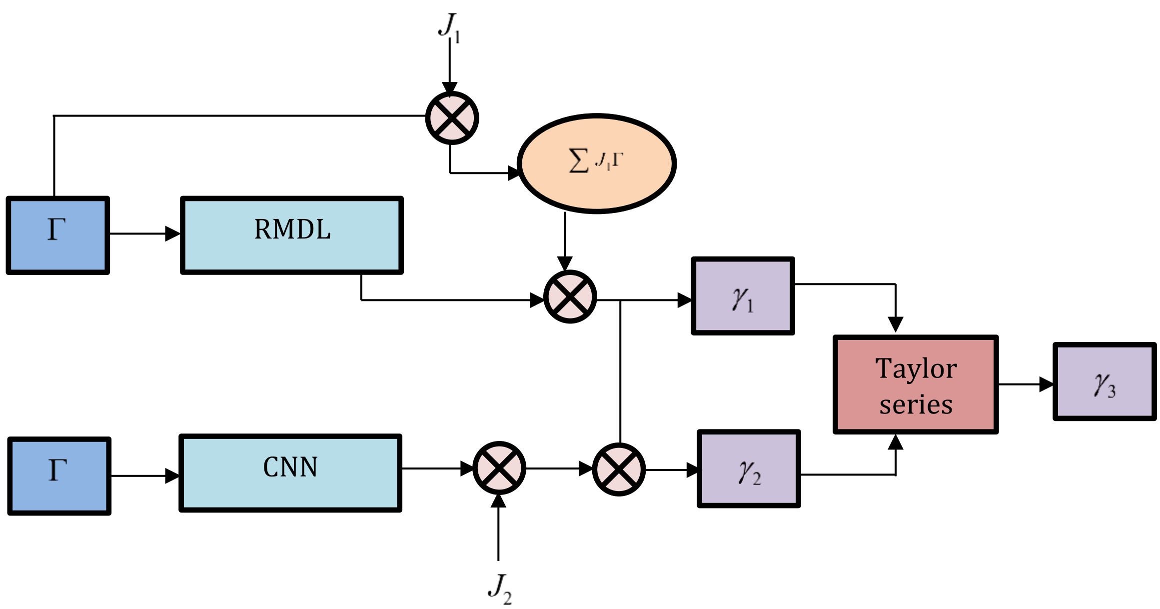


Fig. 2 -Structural representation of RMConv FT-Net for personality prediction

3.5.1. RMDL model

The outcome Γ from imbalanced data is taken as input for the RMDL approach [19]. RMDL can enhance the accuracy and effectiveness of the model. Moreover, DL networks like CNN, RNN, and Deep Neural Networks (DNN) are employed to obtain this approach. DNN is formed by utilizing the architecture of a simple neural network. In addition, DNN is frequently used for categorization and has several hidden layers. CNN is widely used for image categorization, which is performed by exploiting random feature maps and hidden layers, and RNN is utilized for classifying the text. Moreover, The Gated Recurrent Units (GRU) and LSTM are the two RNN structures employed by RMDL. Here, several layers for the DL multi-technique are utilized randomly, and the outcome of RMDL is written as,

$$P(e_{q1}, e_{q2}, \ldots\ldots, e_{q\varsigma}) = \left[\frac{1}{2} + \frac{\left(\sum_{b'=1}^{\varsigma} r_{qb'}\right)}{\varsigma} - \frac{1}{2}\right] \qquad (15)$$

where, ς is exemplified the number of random models, prediction outcome for model$b'$and data point q is indicated as $r_{qb'}$, and majority voting is employed to determine the final outcome $\hat{e}_q$, which is designated as,

$$\hat{e}_q = \left[\hat{e}_{q1} \ldots\ldots \hat{e}_{b'} \ldots\ldots\ldots \hat{e}_{q\varsigma}\right]^{\iota} \qquad (16)$$

Here, the data point for the model $b'$and detecting document label is implied as $\hat{e}_{qb'}$, and it is expressed as,

$$\beta = arg\max_{E} \left[Soft\,max(\Gamma)\right] \qquad (17)$$

where, β represents the outcome from RMDL. The attained output is multiplied by the normalized outcome. $\sum\sum J_1\Gamma$for obtaining γ_1, and it is expressed as,

$$\gamma_1 = \arg\max_{E} [Soft\,max(\Gamma)] * \Gamma * \sum(J_1 * \Gamma) \quad (18)$$

Later, the output γ_1 is forwarded to the RMConv FT-Net layer, and Fig. 3. portrays the structural view of RMDL [19].

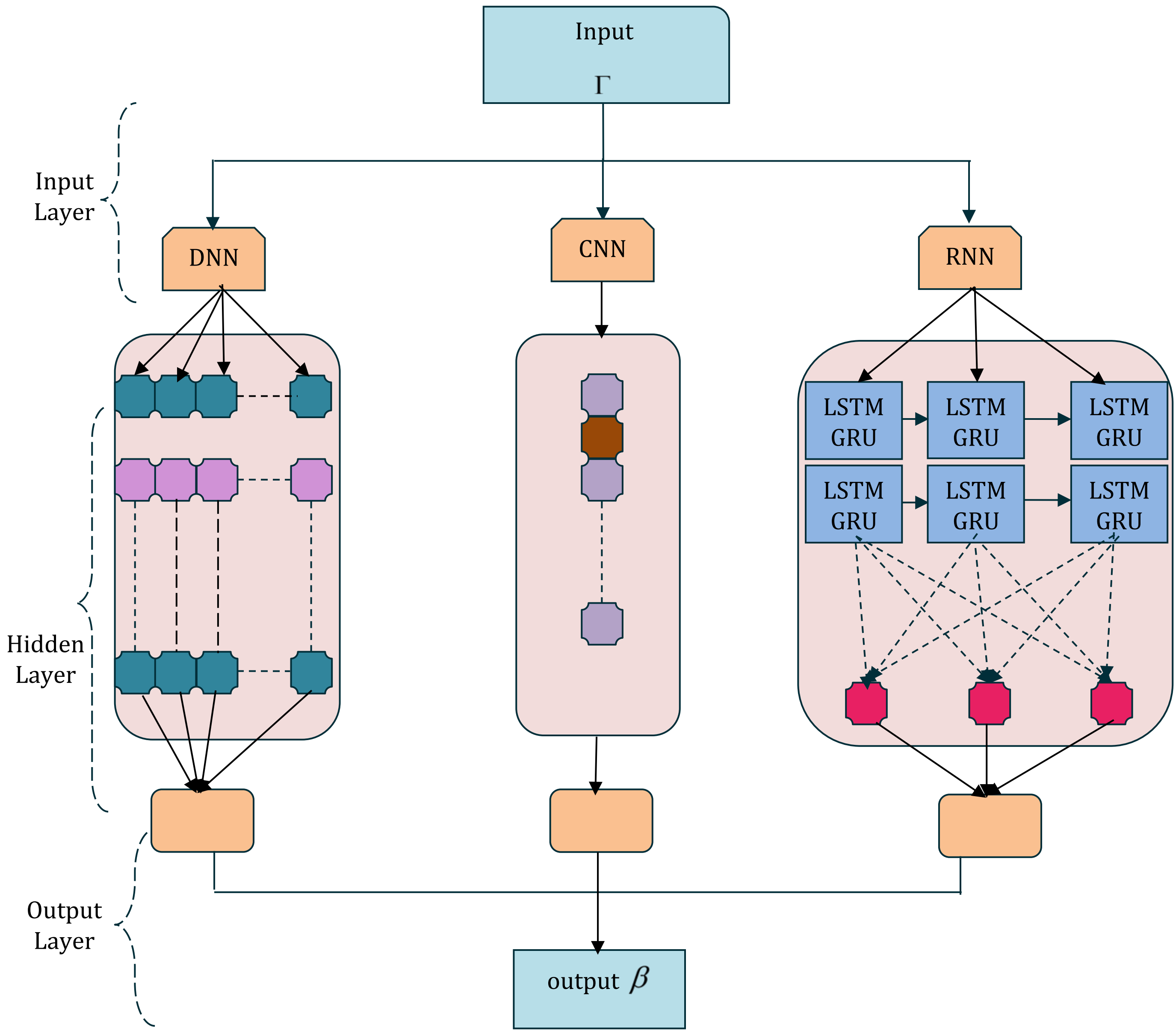


Fig. 3- Architectural diagram of RMDL model.

3.5.2. CNN model

CNN technique [20] is exploited in object detection, pattern recognition, image segmentation, and classification tasks. An appropriate outcome with less computational process is obtained by employing this CNN model. Moreover, the augmented data Γ is considered as input. This technique includes many layers: activation, dense, Batch Normalization (BN), dropout, and Conv2D.

1. Conv layer

This layer is used for mining the features, and the outcome obtained in this layer is passed to the subsequent layer, which includes various interconnected layers. Moreover, the feature value in the acceptance field is transformed into a single value, and it is written as,

$$\varpi = Z^n A \qquad (19)$$

$$Z^n = \left[I_{11}^n, I_{12}^n, \ldots, I_{1u'}^n, \ldots, I_{u'u'}^n\right]^V \in {R_e}^{y^2 \times \varepsilon^2} \qquad (20)$$

$$A = \left[A_1, A_2, \ldots, A_{r'k'}\right] \in {R_e}^{\varepsilon^2 \times r'k'} \qquad (21)$$

$$\varpi = \left[\varpi_1, \varpi_2, \ldots, \varpi_{r'k'}\right] \in {R_e}^{y^2 \times r'k'} \qquad (22)$$

Here, ϖ represent the output feature, Z^n is a column vector with filter weights while A represents the exemplified filter weight and k' is defined as the total sum of channels, in addition to $A_{r'k'}$ embodies weight of $r'k'^{th}$ channel [20].

2. Activation layer

This layer is exploited to gather data regarding the asymmetric relationship between input and outcome. The equation of the layer is designated as,

$$\lambda = \zeta(R) = \zeta(Z^n A) \qquad (23)$$

The activation function is signified as ζ, and λ is implied as the activation layer output.

3. Dropout

A dropout layer is exploited to invalidate neurons' contributions and improve generalization performance. Moreover, dropout improves the training process and minimizes overfitting.

4. Batch Normalization

In this layer, the input distribution is transformed into a typical distribution with variance and mean, which is exploited to create a distribution drop during the activation function phase. The gradient distribution issues are eliminated, which expedites the training process.

5. Dense

A dense layer, signified as a connected layer, alters the array of features into a structured representation. Moreover, it effectively reveals the relation among several variables in the text feature set.

6. Flatten

A flatten is utilized for changing the resultant feature map of the unidimensional matrix and exploited for rearranging the matrix into vector form, which is depicted as,

$$U = \left[Q_1, Q_2 \ldots Q_{r'k'}\right] \qquad (24)$$

$$= \left[F_1 x_1 F_2 x_2 \ldots . F_{r'k'} x_{r'k'}\right] \qquad (25)$$

where, $x_{r'k'}$ is embodied input of $r'k'^{th}$ layer, while $Q_{r'k'}$ is postulated the output of $r'k'^{th}$ layer and flattening operator is articulated as,

$$H = \psi[B] = \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_{r'k'} \end{bmatrix} = \begin{bmatrix} F_1 x_1 \\ F_2 x_2 \\ \vdots \\ T_{r'k'} x_{r'k'} \end{bmatrix} \quad (26)$$

where, F specified the identity matrix, H is stipulated flatten layer output, and ψ is epitomized by the flatten operator. Hence, the outcome is articulated as,

$$P = H_{r'z'}\big(C_{r'z'} \ldots H_2\big(C_2 H_1(C_1 H)\big) \ldots\big) \quad (27)$$

where, z' is implied as a fully connected layer, H is designated activation function, and C embodies weight of the matrix. Moreover, the result of the activation function vector is mentioned as $D \underline{\Delta} H_{r'z'}$.

Hence, the outcome is articulated as,

$$P = D\left(C_{r'z'} \ldots H_1\big(C_1 \psi\{F_{r'\lambda} \zeta_{r'\lambda}([\ldots[F_2 \zeta_2([F_1 \zeta_1(\Gamma)]^{n_2} A_2)] \ldots]^{n_{r'\lambda}} A_{r'\lambda})\}\big)\right) \quad (28)$$

The attained outcome is multiplied by the weight J_2 and multiplied with γ_1 for attaining the output γ_2, and it is written as,

$$\gamma_2 = D\left(C_{r'z'} \ldots H_1\big(C_1 \psi\{F_{r'\lambda} \zeta_{r'\lambda}([\ldots[F_2 \zeta_2([F_1 \zeta_1(\Gamma)]^{n_2} A_2)] \ldots]^{n_{r'\lambda}} A_{r'\lambda})\}\big)\right) * J_2\left(\arg\max_E [Soft\,max(\Gamma)] * \Gamma * \sum(J_1 * \Gamma)\right) \quad (29)$$

Here, Γ indicates input of CNN and Conv is specified as λ, P embodies output from CNN, which is given to the RMConv FT-Net layer as input, γ_2 indicates the product of γ_1 and the weighted output of the CNN, and F shows identity matrix. The architecture of CNN is illustrated in Fig. 4.

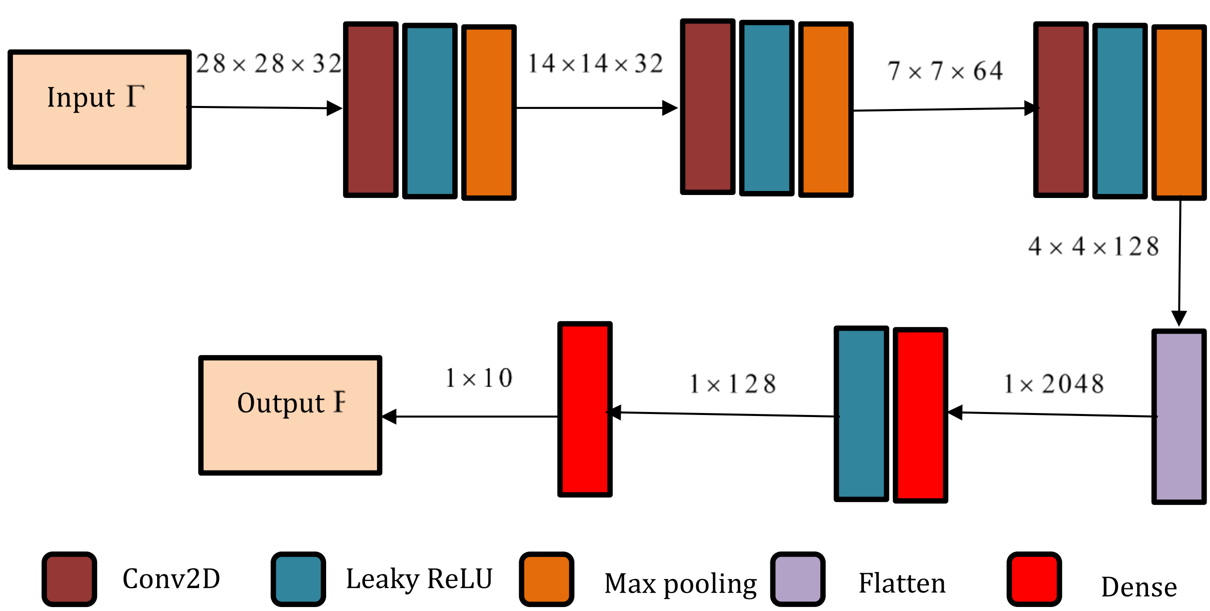


Fig.4 Architecture of CNN

3.5.3. RMConv FT-Net layer

The outcomes γ_1 and γ_2 are given to the RMConv FT-Net layer. Moreover, outputs are integrated by utilizing the Taylor concept [21], and the outcome from the RMConv FT-Net layer is characterized as γ_3. Here, Taylor's concept is used to lessen the cost and to increase the classification accuracy, and it is expressed as,

$$T(v^{'}+1)=T(v^{'})+\frac{T^{'}(v^{'})}{1!} \quad (30)$$

Here,

$$T^{'}(v^{'})=\frac{T(v^{'})-T(v^{'}-m^{'})}{m^{'}} \quad (31)$$

Let $m'=1$ in $T'(v')$, and substituting the value of $T^{'}(v^{'})$ in equation (28)

$$T(v^{'}+1)=T(v^{'})+\frac{T(v^{'})-T(v^{'}-1)}{1!} \quad (32)$$

$$T(v^{'}+1)=T(v^{'})\big[1-T(v^{'}-1)\big] \quad (33)$$

Assume, $T(v')=\gamma_1$, and $T(v'-1)=\gamma_2$,and $T(v^{'}+1)=\gamma_3$, we get

$$\gamma_3=\gamma_1[1-\gamma_2] \quad (34)$$

Substituting the value of γ_1 and γ_2,

$$\gamma_3 = arg\max_{E}\,[Soft\,max(\Gamma)]*\Gamma*\sum(J_1*\Gamma)$$
$$\Bigg[1-\Bigg(D\bigg(C_{r^{'}z^{'}}\ldots H_1\big(C_1\psi\{F_{r^{'}\lambda}\zeta_{r^{'}\lambda}([\ldots[F_2\zeta_2([F_1\zeta_1(\Gamma)]^{n_2}A_2)]\ldots]^{n_{r^{'}\lambda}}A_{r^{'}\lambda})\}\big)\bigg)*$$
$$J_2\bigg(arg\max_{E}\,[Soft\,max(\Gamma)]*\Gamma*\sum(J_1*\Gamma)\bigg)\Bigg)\Bigg] \quad (35)$$

Here, the output of the RMConv FT-Net layer is specified as γ_3, which is the personality-predicted output [21].

4. Results and discussion

This section details the explanation of the RMConv FT-Net model based on the performance metrics through conventional personality prediction techniques.

4.1. Experimental set-up

In this case, an introduced RMConv FT-Net technique is implemented using Python 3.11. The experiment uses a PC with Intel Core i7, CPU 2.7 GHz, and 8 GB RAM.

4.2. Dataset description

The MBTI dataset [26] is a personality type system that divides everyone into 16 distinct personality types across 4 axes, such as Introversion (I) – Extroversion (E), Intuition (N) – Sensing (S), Thinking (T) – Feeling (F), and Judging (J) – Perceiving (P). This database comprises over 8600 rows of data; on every row is a person's Type (4-letter MBTI code/type). A segment of every of the final 50 things they have posted (Every entry divided by "|||" (3 pipe characters)).

4.3. Performance metrics

The proposed RMConv FT-Net technique is assessed using MSE, MAPE, and RMSE measures.

1. MSE

MSE signifies the average squared difference between expected and acquired values by the RMConv FT-Net and is articulated in equation (36)

$$MSE = \frac{1}{N}\sum_{\hbar=1}^{N}[\gamma_{3\hbar}^{*} - \gamma_{3\hbar}]^2 \quad (36)$$

Here, the actual outcome is denoted as $\gamma_{3\hbar}^{*}$, Output from the RMConv FT-Net model is denoted by γ_3, The total number of samples is represented by N [27].

2. *MAPE*

MAPE, also called Mean Absolute Percentage Deviation (MAPD), is a parameter of prediction accuracy of a personality prediction model, and the expression is stated in equation (37),

$$MAPE = \frac{1}{N}\sum_{\hbar=1}^{N}\left|\frac{\gamma_{3}{}_{\hbar}^{*} - \gamma_{3\hbar}}{\gamma_{3}{}_{\hbar}^{*}}\right| \quad (37)$$

Where $\gamma_{3\hbar}^{*}$ represents the actual outcome, while the γ_3 is the RMConv FT-Net model output, and N is the total number of samples [28].

3. RMSE

The square root of the squared error between the actual and expected samples is also known as RMSE. RMSE is computed using the equation below.

$$RMSE = \sqrt{\frac{1}{N}\sum_{\hbar=1}^{N}[\gamma_{3\hbar}^{*} - \gamma_{3\hbar}]^2} \quad (38)$$

$\gamma_{3\hbar}^{*}$ is the actual value outcome value, and the predicted value from the RMConv FT-Net model is denoted as γ_3 while the N represents the total number of dataset samples [29].

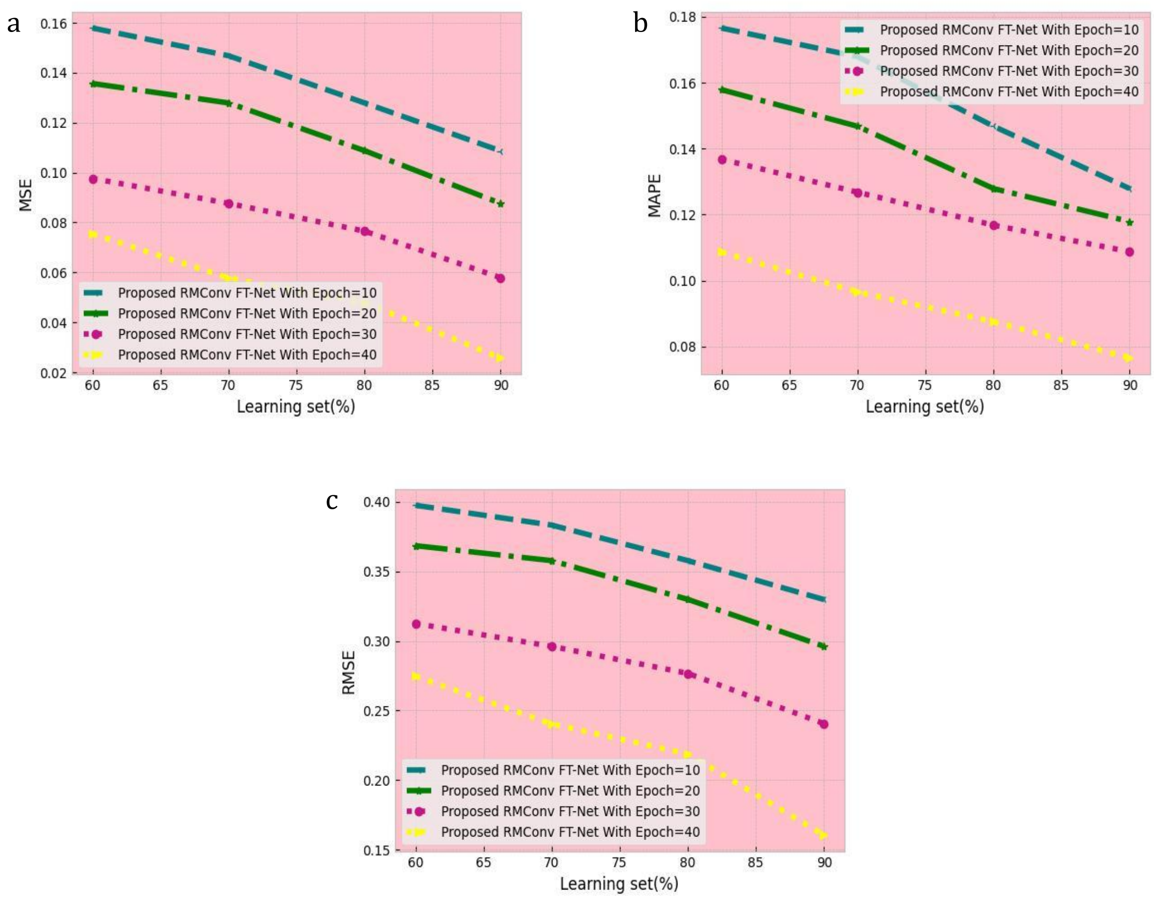
4.4. Performance assessment

The RMConv FT-Net framework is evaluated for its efficiency based on the k-value and learning set.

4.4.1. Analysis based on learning set

The RMConv FT-Net model is assessed for efficiency by contemplating a learning set with changing epochs regarding several metrics. The performance of RMConv FT-Net is examined utilizing a learning set depicted in Fig. 5. The analysis of RMConv FT-Net employing MSE is portrayed in Fig. 5(a). The MSE acquired by RMConv FT-Net employing a learning set as 90% at changing numerous epochs like 10, 20, 30, and 40 is 0.109, 0.088, 0.058, and 0.026. Furthermore, the evaluation of RMConv FT-Net concerning MAPE is signified in Fig. 5(b). Here, the MAPE recorded by RMConv FT-Net utilizing learning set as 90% for ten epochs is 0.128, for 20 epochs is 0.118, for 30 epochs is 0.109, and for 40 epochs is 0.077. Moreover, Fig. 5(c) displays an assessment of RMConv FT-Net concerning RMSE. With a 90% learning set, the RMSE measured by RMConv FT-Net is 0.330, 0.296, 0.241, and 0.161 at 10, 20, 30, and 40 epochs.

We compared the proposed model with previous studies for a more accurate evaluation depending on the learning set value. The estimation of RMConv FT-Net employing a learning set is displayed in Fig. 6. Fig. 6(a) portrays the evaluation of RMConv FT-Net concerning MSE. The MSE acquired with learning set as 90% is 0.309, 0.248, 0.209, 0.109, and 0.026, respectively, for DL-based fusion, LR+SMOTE, Stacked ensemble framework, and MFCC-based NPSO_CNN, and RMConv FT-Net. Likewise, Fig. 6(b) depicts the assessment of RMConv FT-Net based on MAPE. In



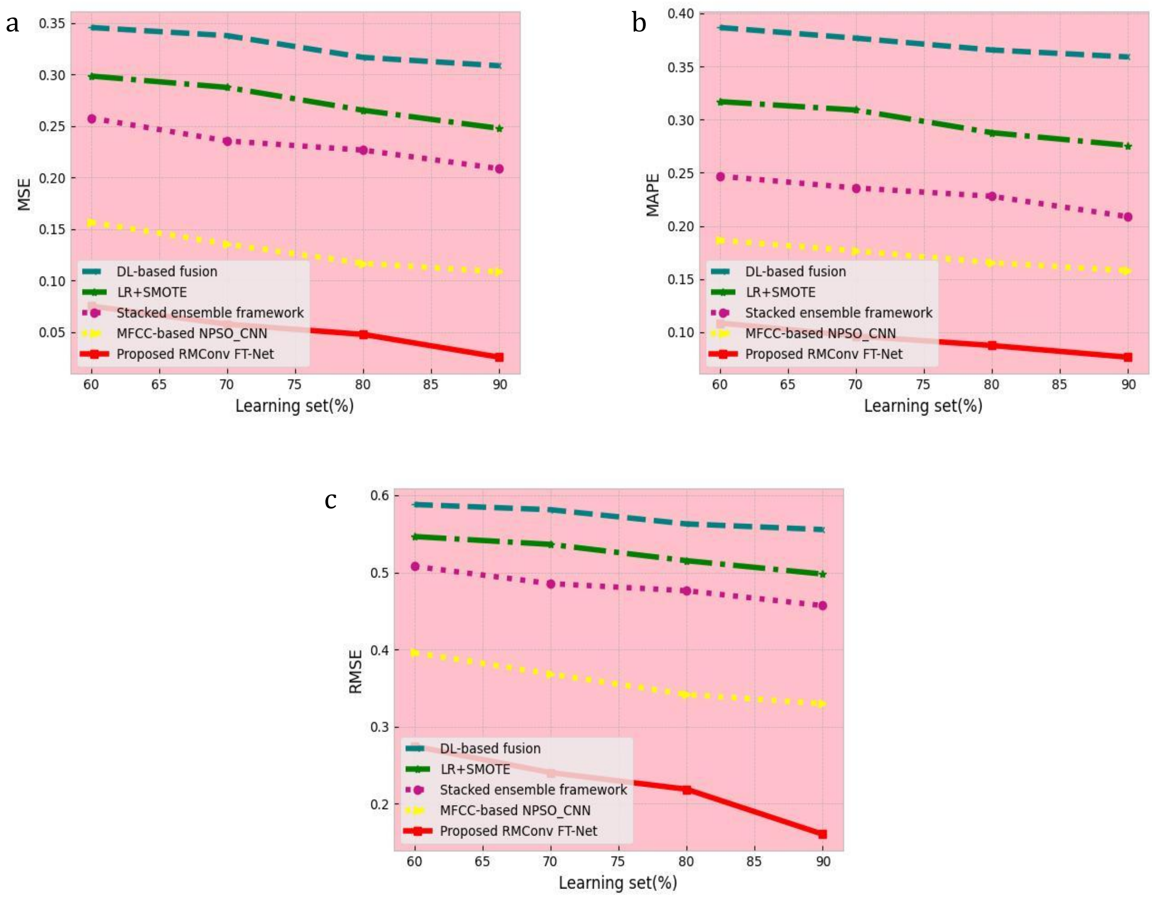
this case, RMConv FT-Net has recorded MAPE of 0.077, and prevailing schemes like DL-based fusion, LR+SMOTE, stacked ensemble framework, and MFCC-based NPSO_CNN measured MAPE of 0.359, 0.276, 0.209, and 0.158, with learning set as 90%. Fig. 6(c) signifies the analysis of RMConv FT-Net to RMSE. The RMConv FT-Net quantified an RMSE of 0.161, and the conventional techniques, including DL-based fusion, LR+SMOTE, stacked ensemble framework, and MFCC-based NPSO_CNN, have quantified an RMSE of 0.556, 0.498, 0.457, and 0.330, with learning set as 90%.

Fig. 5- Analysis of RMConv FT-Net technique employing (a) MSE; (b) MAPE; (c) RMSE with learning set.

4.4.2 Analysis based on k-value

Fig. 7. specifies the performance assessment of RMConv FT-Net by changing the k-value. In Fig. 7(a), the evaluation of RMConv FT-Net concerning MSE is illustrated. If the K-value is 8, the devised RMConv FT-Net model obtained MSE of 0.075, 0.065, 0.058, and 0.017 with epoch 10, 20, 30, and 40. Fig. 7(b) indicates the RMConv FT-Net method concerning MAPE. The proposed RMConv FT-Net acquired MAPE with epoch ten is 0.109, 20 is 0.098, 30 is 0.088, and 40 is 0.047 by employing the k-value as 8. In Fig. 7(c), the performance valuation using MAPE is depicted. Here, RMSE acquired by RMConv FT-Net is 0.275 with epoch 10, 0.265 with epoch 20, 0.241 with epoch 30, and 0.129 with epoch 40, while considering k-value is 8.

Additionally, the proposed model has been compared according to K-values with other previous models, and the results of the comparative analysis are illustrated in Fig. 8. Fig 8(a) demonstrates the investigation of RMConv FT-



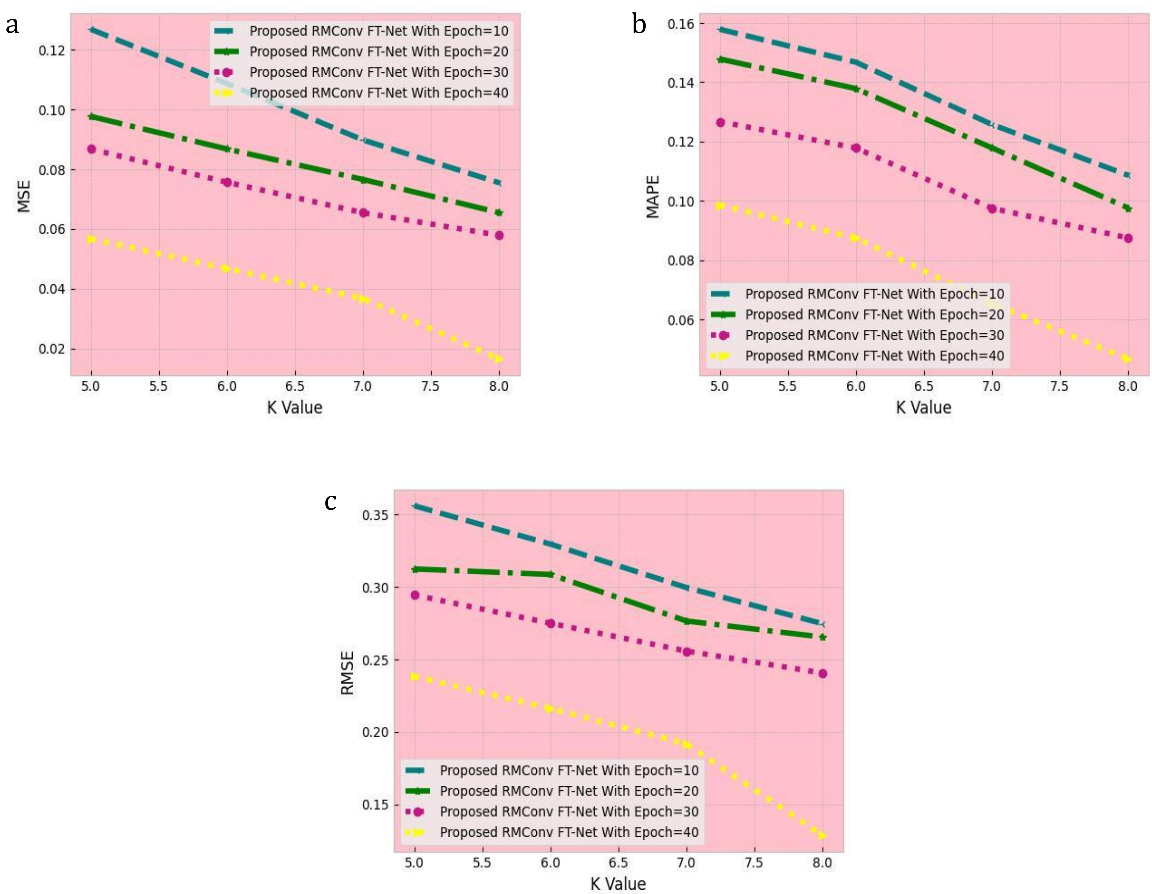
Net employing MSE with k-value. The MSE acquired utilizing k-value as eight is 0.258, 0.165, 0.109, 0.047, and 0.017 for DL-based fusion, LR+SMOTE, stacked ensemble framework, and MFCC-based NPSO_CNN, and RMConv FT-Net. In addition, Fig. 8(b) represents the estimation of RMConv FT-Net employing MAPE. The RMConv FT-Net has achieved a MAPE of 0.047, and prevailing techniques like DL-based fusion, LR+SMOTE, stacked ensemble framework, and MFCC-based NPSO_CNN obtained a MAPE of 0.309, 0.258, 0.165, and 0.127 with k-value as 8. Furthermore, Fig. 8(c) signifies the evaluation of RMConv FT-Net concerning RMSE. With a k-value of 8, the RMSE attained is 0.508 for DL-based fusion, 0.407 for LR+SMOTE, 0.330 for stacked ensemble framework, 0.216 for MFCC-based NPSO_CNN, and 0.129 for RMConv FT-Net.

Fig. 6- Assessment of RMConv FT-Net utilizing (a) MSE; (b) MAPE; and (c) RMSE with learning set.

4.5 Comparative discussion

The RMConv FT-Net is proposed to employ assessment metrics, including MSE, MAPE, and RMSE, as portrayed in Table 1. Table 1 illustrates the MSE, MAPE, and RMSE of RMConv FT-Net compared with conventional methods with a k-value of 8 and a learning set of 90%. DL-based fusion, LR+SMOTE, stacked ensemble framework, and MFCC-based NPSO_CNN quantify the MSE value of 0.258, 0.165, 0.109, and 0.047. The prevailing techniques like DL-based fusion, LR+SMOTE, stacked ensemble framework, and MFCC-based NPSO_CNN obtained a MAPE of 0.309, 0.258, 0.165, and 0.127. The RMSE attained is 0.508 for DL-based fusion, 0.407 for LR+SMOTE, 0.330 for stacked ensemble framework, and 0.216 for MFCC-based NPSO_CNN. Incorporating RMDL and CNN in the RMConv FT-Net model led to the proposed model accumulating the accuracy and high robustness of the RMDL with the minimal computational complexity of the CNN, thus resulting in superior performance in predicting personality.

Even though the training set was 90% of the data, the results are reliable for many reasons. First, using k-fold cross-validation enables the system to learn from various parts of data with different contexts representing personality. This process is conducted by dividing the data into k-folds and then training the model on the K-1 fold to evaluate it on the remaining fold. Accordingly, repeating this process of training K epochs, relying on various folds for evaluation each time, helps to provide a more reliable estimate of the model's performance. Second, 90/10 data splitting represents a reasonable computational balance between efficiency and computational cost since a small



test set needs fewer computational resources for the evaluation step.

Fig. 7- Assessment of RMConv FT-Net model employing (a) MSE; (b) MAPE; (c) RMSE with k-value.

The proposed model has a computational complexity that can cause slow training convergence. Additionally, this framework does not integrate transfer models, which could help capture contextual information from the text. Moreover, the model's application is restricted to personality traits without extending to other applications such as sentiment analysis and emotions.

Table 1. Comparative discussion of RMConv FT-Net model.

Variations	Metrics	stacked ensemble framework [5]	DL-based fusion [11]	LR+ SMOTE [12]	MFCC-based NPSO_CNN [13]	Proposed RMConv FT-Net
Learning set as 90	MSE	0.209	0.309	0.248	0.109	0.026
	MAPE	0.209	0.359	0.276	0.158	0.077

	RMSE	0.457	0.556	0.498	0.330	**0.161**
K-value as 8	**MSE**	0.109	0.258	0.165	0.047	**0.017**
	MAPE	0.165	0.309	0.258	0.127	0.047
	RMSE	0.330	0.508	0.407	0.216	0.129

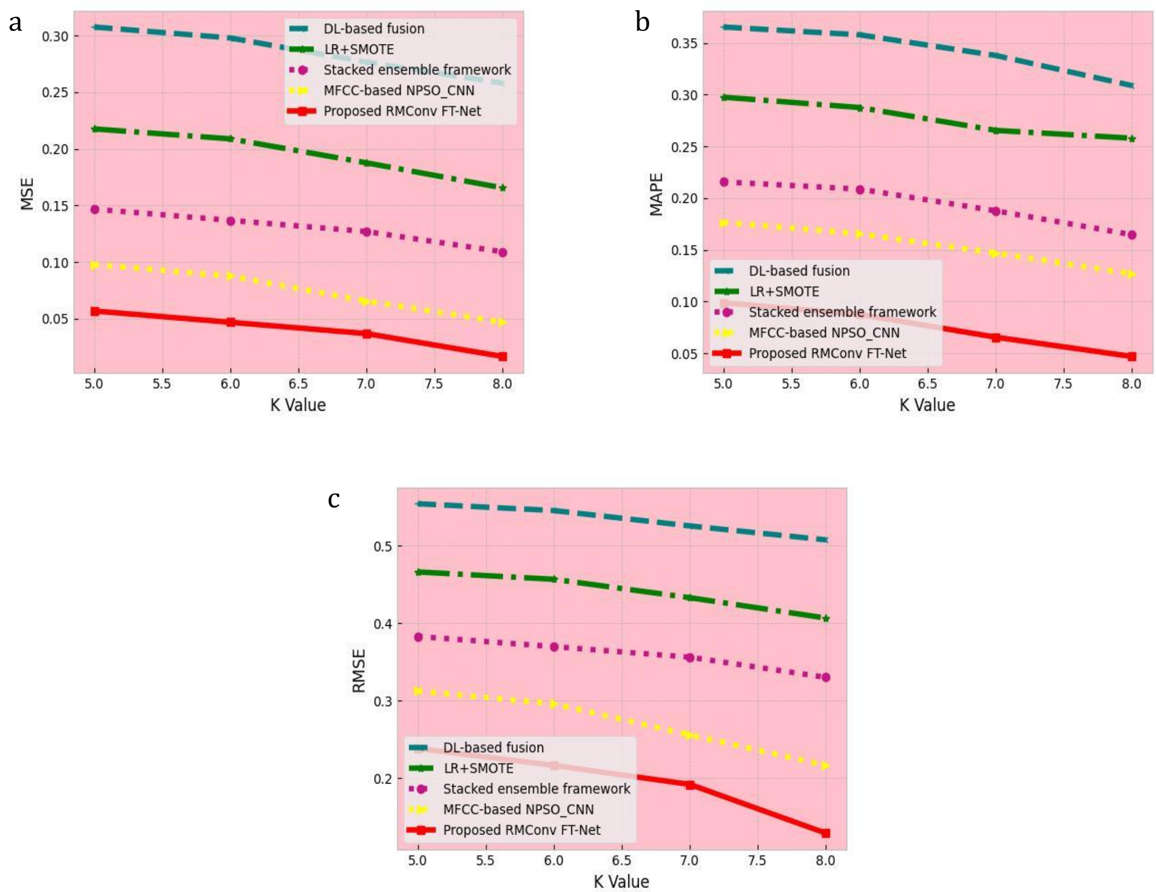


Fig. 8- Investigation of RMConv FT-Net employing (a) MSE; (b) MAPE; (c) RMSE with k-value.

5. Conclusion

Personality prediction is employed to understand a person's traits and is mainly performed based on text data. Personality significantly influences many aspects of human lives, including desires, mental health, and life decisions. Therefore, a new method known as RMConv FT-Net has been devised to predict personality. Initially, text data from the database is considered input, and then tokenization is performed to split the texts into tokens, which BERT does. After this, various features, including linguistic features and several other features, like TF, TF-IDF, Word2Vec, and so on, are mined from a text. Following this, data imbalance is handled by employing oversampling. At last, personality prediction is performed by utilizing RMConv FT-Net, which is formed by incorporating CNN, Taylor series, and RMDL. The experimental outcomes of RMConv FT-Net based on MSE, MAPE, and RMSE show that it recorded superior values of 0.017, 0.047, and 0.129, respectively. The proposed model merges the features of CNN and DL into one system with optimization to reach a reasonable optimality in predicting personality. Even though the proposed model used various features such as TF, TF-IDF, and Word2Ve, the model does not use more advanced features from transfer learning or pre-learned models that could improve the accuracy. This limitation suggests that the current feature set may need to capture the full complexity of personality traits. Additionally, the model of RMConv FT-Net was trained and adjusted for the purpose of personality prediction as the model may not be generalizable across different populations or contexts. Future work aims to enhance the fusion strategy by integrating various word-to-vector models. Moreover, emotion features and several kinds of attention mechanisms, namely self-attention and weighted attention, can also be executed to combine the modalities.

References

[1] S. Fyffe, P. Lee, and S. Kaplan, "Transforming' Personality Scale Development: Illustrating the Potential of State-of-the-Art Natural Language Processing," Organizational Research Methods, vol. 27, (2024), pp. 265–300.

[2] K. R. Chowdhary, "Natural Language Processing," in Fundamentals of Artificial Intelligence, K. R. Chowdhary, Ed., New Delhi: Springer India, 2020, pp. 603–649.

[3] Y. Mehta, N. Majumder, A. Gelbukh, and E. Cambria, "Recent Trends in Deep Learning Based Personality Detection," Artif Intell Rev, vol. 53, (2020), pp. 2313–2339.

[4] J. Jang, S. Yoon, G. Son, M. Kang, J. Y. Choeh, and K.-H. Choi, "Predicting Personality and Psychological Distress Using Natural Language Processing: A Study Protocol," Front. Psychol., vol. 13, (2022).

[5] J. Serrano-Guerrero, B. Alshouha, M. Bani-Doumi, F. Chiclana, F. P. Romero, and J. A. Olivas, "Combining machine learning algorithms for personality trait prediction," Egyptian Informatics Journal, vol. 25, (2024), p. 100439.

[6] A. Bhagat, A. Aylani, A. Shaahsane, S. Shinde, and H. Limaye, "Personality Prediction By Using Deep Learning On Twitter." Rochester, NY, (2023).

[7] Z. Reisz, M. J. Boudreaux, and D. J. Ozer, "Personality traits and the prediction of personal goals," Personality and Individual Differences, vol. 55, (2013), pp. 699–704.

[8] M. A. Kosan, H. Karacan, and B. A. Urgen, "Predicting personality traits with semantic structures and LSTM-based neural networks," Alexandria Engineering Journal, vol. 61, (2022), pp. 8007–8025.

[9] D. Xue et al., "Deep learning-based personality recognition from text posts of online social networks," Appl Intell, vol. 48, (2018), pp. 4232–4246.

[10] H. Ahmad, M. U. Asghar, M. Z. Asghar, A. Khan, and A. H. Mousavi, "A Hybrid Deep Learning Technique for Personality Trait Classification From Text," IEEE Access, vol. 9, (2021), pp. 146214–146232.

[11] K. El-Demerdash, R. A. El-Khoribi, M. A. Ismail Shoman, and S. Abdou, "Deep learning based fusion strategies for personality prediction," Egyptian Informatics Journal, vol. 23, (2022), pp. 47–53.

[12] G. Ryan, P. Katarina, and D. Suhartono, "MBTI Personality Prediction Using Machine Learning and SMOTE for Balancing Data Based on Statement Sentences," Information, vol. 14, (2023), p. 217.

[13] K. Rangra, V. Kadyan, and M. Kapoor, "Emotional speech-based personality prediction using NPSO architecture in deep learning," Measurement: Sensors, vol. 25,(2023), p. 100655.

[14] M. Ramezani, M.-R. Feizi-Derakhshi, and M.-A. Balafar, "Knowledge Graph-Enabled Text-Based Automatic Personality Prediction," Computational Intelligence and Neuroscience, vol. 2022, (2022), p. 3732351.

[15] M. Ramezani et al., "Automatic personality prediction: an enhanced method using ensemble modeling," Neural Comput & Applic, vol. 34, (2022), pp. 18369–18389.

[16] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv, (2019).

[17] R. K. Thakur and M. V. Deshpande, "Kernel Optimized-Support Vector Machine and MapReduce Framework for Sentiment Classification of Train Reviews," Int. J. Unc. Fuzz. Knowl. Based Syst., vol. 27, (2019), pp. 1025–1050.

[18] B. Ohana and B. Tierney, "Sentiment Classification of Reviews Using SentiWordNet," Conference papers,(2009).

[19] K. Kowsari, M. Heidarysafa, D. E. Brown, K. J. Meimandi, and L. E. Barnes, "RMDL: Random Multimodel Deep Learning for Classification," in Proceedings of the 2nd International Conference on Information System and Data Mining, in ICISDM '18. New York, NY, USA: Association for Computing Machinery, pp. 19–28, (2024).

[20] H.-T. Nguyen, S. Li, and C. C. Cheah, "A Layer-Wise Theoretical Framework for Deep Learning of Convolutional Neural Networks," IEEE Access, vol. 10, (2022), pp. 14270–14287.

[21] S. A. Mangai, B. R. Sankar, and K. Alagarsamy, "Taylor Series Prediction of Time Series Data with Error Propagated by Artificial Neural Network," International Journal of Computer Applications, vol. 89, (2014), pp. 41–47.

[22] A. Jalilifard, V. F. Caridá, A. F. Mansano, R. S. Cristo, and F. P. C. da Fonseca, Semantic Sensitive TF-IDF to Determine Word Relevance in Documents, vol. 735, (2021).

[23] Y. Ren and D. Ji, "Learning to Detect Deceptive Opinion Spam: A Survey," IEEE Access, vol. 7, (2019), pp. 42934–42945.

[24] S. K. Saha, S. Sarkar, and P. Mitra, "A Hybrid Feature Set based Maximum Entropy Hindi Named Entity Recognition," in Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I, (2008).

[25] A. Gupta, K. Sharma, and K. K. Goyal, "Measuring similarity between verb pairs in the biomedical domain: An ontology-based information content perspective," International journal of health sciences, vol. 6, (2022), pp. 5099–5112.

[26] "(MBTI) Myers-Briggs Personality Type Dataset." Accessed: Aug. 05, 2024. [Online]. Available: https://www.kaggle.com/datasets/datasnaek/mbti-type.

[27] B. H. Nayef, S. N. H. S. Abdullah, and M. Mohammed, "Deep Learning Function Implications For Handwritten Character Recognition Using Improved Mean Square Error Function," Baghdad Science Journal, vol. 4, (2024).

[28] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," International Journal of Forecasting, vol. 32, (2016), pp. 669–679.

[29] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?," Geoscientific Model Development Discussions, vol. 7, (2014), pp. 1525–1534.