

HYBRID DEEP LEARNING APPROACH FOR PERSONALITY TRAIT CLASSIFICATION FROM TEXTUAL DATA

¹D. Dada Kalandar, ²Chakali Rakesh, ³Pillala Charan Kumar, ⁴Dasari Mahesh, ⁵Talari Vinod Kumar

¹Assistant Professor, ^{2,3,4,5}Students

Department of Computer Science and Engineering

St. John's College of Engineering & Technology, Yerrakota, Yemmiganur, Kurnool, A.P.

khalandar0509@gmail.com, chakalirakesh3465@gmail.com, charanram2143@gmail.com, armymahesh381@gmail.com, talarivinodkumar014@gmail.com

ABSTRACT

Personality trait detection from textual data has become an important research area in cognitive computing, natural language processing, and social media analytics. With the rapid growth of social networking platforms such as Twitter, Facebook, and Instagram, users generate massive amounts of textual content that reflects their opinions, emotions, and behavioral patterns. These textual expressions provide valuable insights into an individual's personality traits, which can be useful for applications such as recruitment systems, recommendation engines, marketing analysis, and psychological studies. However, accurately identifying personality traits from social media text remains challenging due to the informal nature, diversity, and contextual complexity of online content.

To address these challenges, this study proposes a Hybrid Deep Learning Approach for Personality Trait Classification from Textual Data that integrates Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. The CNN model is employed to capture significant local textual features through convolution and pooling operations, while the LSTM model preserves contextual and sequential information from the text to better understand relationships between words. By combining these two deep learning techniques, the proposed hybrid architecture effectively extracts semantic and contextual features from textual data, enabling improved classification of personality traits such as Introversion–Extroversion, Intuition–Sensing, Thinking–Feeling, and Judging–Perceiving. Experimental evaluations demonstrate that the hybrid CNN–LSTM model achieves higher accuracy and better performance compared to traditional machine learning algorithms and single deep learning models. The proposed system provides an efficient framework for automatic

personality trait classification and can support various real-world applications in social media analysis, human behavior understanding, and intelligent decision-making systems.

Keywords: Personality Trait Classification, Hybrid Deep Learning, CNN–LSTM, Natural Language Processing, Text Mining, Social Media Analytics, Cognitive Computing.

I. INTRODUCTION

Cognitive science is a multidisciplinary research field that focuses on understanding human mental processes such as learning, thinking, perception, memory, emotions, and personality. Among these cognitive processes, personality plays an important role in describing individual behavior, preferences, and social interactions. Personality traits influence how people communicate, make decisions, and express their opinions in different situations. In recent years, researchers have shown significant interest in automatically identifying personality traits using computational methods and artificial intelligence techniques.

With the rapid growth of social media platforms such as Twitter, Facebook, Instagram, and online discussion forums, people generate large amounts of textual data in the form of posts, comments, reviews, and messages. These textual expressions often reflect users' attitudes, emotions, and personality characteristics. Analyzing such data provides valuable insights into human behavior and can help in understanding psychological patterns, consumer behavior, and user preferences. As a result, automatic personality trait detection from textual data has become an important research topic in areas such as natural language processing, cognitive computing, and sentiment analysis.

Traditional methods for personality assessment typically rely on psychological questionnaires and manual

evaluation processes conducted by experts. Although these methods can provide reliable results, they are time-consuming, expensive, and not suitable for analyzing large populations or massive social media data. To overcome these limitations, machine learning and data-driven approaches have been introduced to automatically analyze textual content and classify personality traits. Early studies mainly used conventional machine learning algorithms such as Support Vector Machines (SVM), Decision Trees, and Random Forests for personality classification tasks. However, these methods often rely on handcrafted features and simple text representations, which may not effectively capture the complex semantic and contextual relationships present in natural language. Recent advancements in deep learning and natural language processing have provided powerful tools for analyzing textual data more effectively. Deep learning models such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) networks have demonstrated strong capabilities in extracting meaningful patterns from large datasets. CNN models are effective in capturing local textual features and patterns, while LSTM networks are capable of preserving contextual information and learning long-term dependencies between words in a sequence.

To improve the performance of personality trait classification, researchers have proposed hybrid deep learning architectures that combine the strengths of multiple models. In this research, a Hybrid Deep Learning Approach using CNN and LSTM is proposed for personality trait classification from textual data. The CNN component extracts important local features from text, while the LSTM layer captures contextual relationships and sequential dependencies among words. This combined architecture enhances the system's ability to accurately classify personality traits such as Introversion–Extroversion, Intuition–Sensing, Thinking–Feeling, and Judging–Perceiving based on textual content.

The proposed hybrid deep learning framework aims to improve the accuracy and efficiency of automatic personality detection systems. Such systems can be applied in various domains including social media analysis, recommendation systems, recruitment processes, customer behavior analysis, and

psychological research. By leveraging advanced deep learning techniques, the system provides a scalable and effective solution for understanding personality traits from large volumes of textual data.

II. LITERATURE SURVEY

Personality trait classification from textual data has attracted significant research attention in the fields of natural language processing, cognitive computing, and social media analytics. Researchers have explored various machine learning and deep learning approaches to automatically identify personality traits from user-generated textual content such as social media posts, blogs, and online reviews.

Early studies mainly focused on supervised machine learning techniques for personality recognition. These approaches use labeled datasets to train classification models that predict personality traits based on linguistic features extracted from textual data. Traditional algorithms such as Support Vector Machines (SVM), Decision Trees, Logistic Regression, and Random Forests have been widely applied for personality detection tasks. These models analyze features such as word frequency, syntactic patterns, and linguistic styles to classify personality traits. Although supervised machine learning methods provide reasonable performance, they often depend heavily on manually engineered features and annotated datasets, which limits their scalability and efficiency.

In contrast, unsupervised machine learning approaches have been explored when labeled data is limited or unavailable. These methods classify textual data based on patterns discovered automatically from the data itself. In many studies, keyword-based analysis and clustering techniques have been used to identify personality characteristics from social media content. Such systems typically analyze linguistic patterns and vocabulary usage to estimate personality scores. While unsupervised methods can process large datasets without requiring manual annotations, they often suffer from lower classification accuracy due to the lack of explicit training labels.

Researchers have also investigated hybrid and semi-supervised approaches that combine features from both supervised and lexicon-based methods. These techniques integrate labeled datasets with linguistic dictionaries or rule-based systems to improve personality classification

accuracy. Hybrid systems aim to leverage the strengths of multiple approaches, enabling better feature representation and improved classification performance. With the advancement of artificial intelligence, deep learning techniques have gained popularity for personality trait detection tasks. Deep learning models can automatically learn hierarchical feature representations from large textual datasets without relying on handcrafted features. Convolutional Neural Networks (CNN) have been widely used for text classification because of their ability to capture local textual patterns and semantic relationships within sentences. Similarly, Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks are capable of capturing sequential dependencies and contextual information from text data.

Recent research has demonstrated that combining multiple deep learning models can significantly improve the performance of personality classification systems. Hybrid deep learning architectures that integrate CNN and LSTM models are particularly effective because they combine the strengths of both techniques. The CNN component extracts important local features and semantic representations from textual data, while the LSTM component preserves contextual information and captures long-term dependencies between words. This hybrid architecture allows the model to understand both the structure and context of text more effectively.

Despite these advancements, several challenges remain in personality trait classification from textual data. Social media content often contains informal language, abbreviations, emojis, and inconsistent grammatical structures, which makes accurate text analysis difficult. Additionally, many existing models struggle to capture deeper contextual information and semantic relationships in complex sentences.

Therefore, there is a need for more advanced hybrid deep learning frameworks that can effectively combine feature extraction and contextual learning capabilities. The proposed Hybrid CNN–LSTM model addresses these limitations by integrating convolutional feature extraction with sequential contextual learning, enabling more accurate and efficient personality trait classification from textual data.

III. SYSTEM ANALYSIS EXISTING SYSTEM

Existing systems for personality trait classification from textual data mainly rely on traditional machine learning algorithms such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forest, and Decision Trees. These systems analyze textual data collected from social media platforms and classify personality traits based on predefined linguistic features such as word frequency, keywords, and syntactic patterns.

In many approaches, the textual data is converted into numerical representations using techniques such as Bag-of-Words or TF-IDF. These feature representations are then used to train machine learning classifiers that attempt to predict personality traits like Introversion–Extroversion, Intuition–Sensing, Thinking–Feeling, and Judging–Perceiving. Although these approaches can process large datasets, they often fail to capture deeper contextual relationships between words and sentences.

Furthermore, many existing systems rely on single deep learning models such as CNN or RNN, which may focus only on limited aspects of textual information. For example, CNN models primarily extract local features from text but may ignore sequential context, while some models struggle to handle long-range dependencies in textual data. As a result, the performance of these systems may not be sufficient for accurate personality trait classification.

Disadvantages of Existing System

1. **Lower Classification Accuracy**
Traditional machine learning algorithms often rely on simple text features and cannot effectively capture complex semantic relationships in textual data.
2. **Limited Context Understanding**
Single-model approaches such as CNN focus mainly on local features and fail to preserve contextual information across long text sequences.
3. **Dependence on Manual Feature Engineering**
Many existing systems require handcrafted features and predefined linguistic patterns, which may not generalize well across different datasets and domains.

PROPOSED SYSTEM

The proposed system introduces a **Hybrid Deep Learning Approach for Personality Trait**

Classification from Textual Data by combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. In this approach, the CNN model is first used to extract important local textual features through convolution and pooling operations. These extracted features represent significant patterns within sentences and phrases.

After feature extraction, the extracted features are passed to the LSTM network. The LSTM model analyzes sequential relationships between words and preserves contextual information over long text sequences. This combination allows the system to capture both local semantic patterns and long-term contextual dependencies in textual data.

By integrating the strengths of CNN and LSTM models, the proposed hybrid architecture provides more accurate personality trait classification. The system is capable of efficiently analyzing large volumes of social media text and identifying personality traits such as Introversion–Extroversion, Intuition–Sensing, Thinking–Feeling, and Judging–Perceiving.

Advantages of Proposed System

1. **Higher Classification Accuracy**
The hybrid CNN–LSTM model improves prediction accuracy by combining feature extraction and contextual learning capabilities.
2. **Better Contextual Understanding of Text**
The LSTM component preserves sequential and contextual information, enabling more accurate interpretation of textual data.
3. **Automatic Feature Learning**
The deep learning architecture automatically learns important features from textual data without requiring manual feature engineering.

IV. MODULES

To implement this project we have designed following modules

Upload Personality Trait Dataset: using this module we will upload dataset to application

Pre-process Dataset: using this module we will read all dataset values and then remove stop words and special symbols from all posts

Convert Text to TF-IDF Vector: using this module we will convert all words in a review to vector and this vector contains average frequency of each words

Run CNN+LSTM Algorithm: using this module we will input VECTOR to CNN to trained a model and then extract features from and this features will be input LSTM model for retraining and then perform prediction on test data and calculate it's accuracy

Run SVM Algorithm: using this module we will train SVM algorithm

Run KNN Algorithm: using this module we will train KNN algorithm

Run Random Forest Algorithm: using this module we will train Random forest and then calculate its prediction accuracy

Comparison Graph: using this module we will plot accuracy comparison between all algorithms

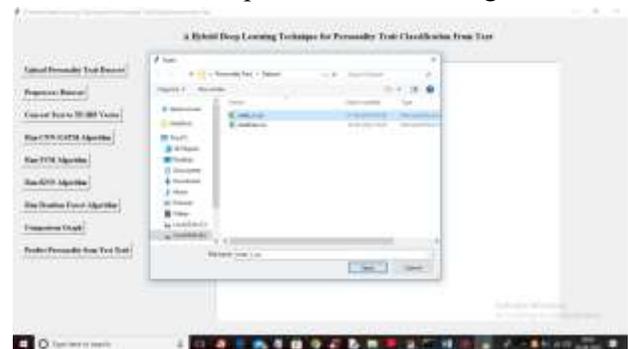
Predict Personality from Test Trait: using this module we will upload test data and then CNN+LSTM will analyse posts from that test data and classify personality trait.

V. RESULTS AND DISCUSSION

To run project double click on 'run.bat' file to get below output



In above screen click on 'Upload Personality Trait Dataset' button to upload dataset and to get below screen



In above screen selecting and uploading dataset file and then click on 'Open' button to load dataset and to get below output



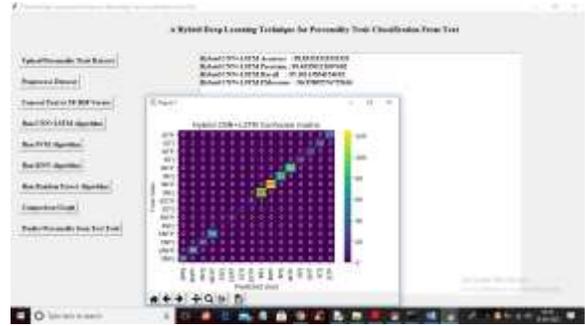
In above screen dataset loaded and now click on 'Preprocess Dataset' button to read and remove all stop words and special symbol from dataset and get below output



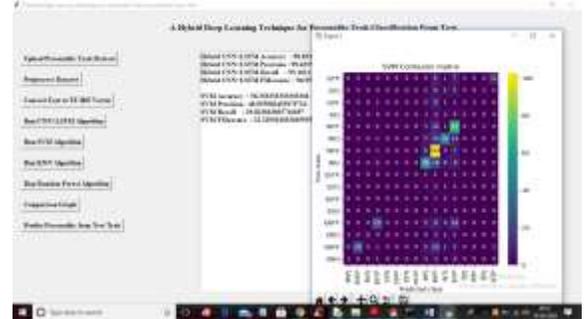
In above screen all posts are cleaned and now click on 'Convert Text to TF-IDF Vector' button to convert above posts messages to TF-IDF vector and get below output



In above screen in first line we can see all words from all posts and remaining lines contains average frequency of those words and now click on 'Run CNN+LSTM Algorithm' button to train above vector with CNN+LSTM and get below output



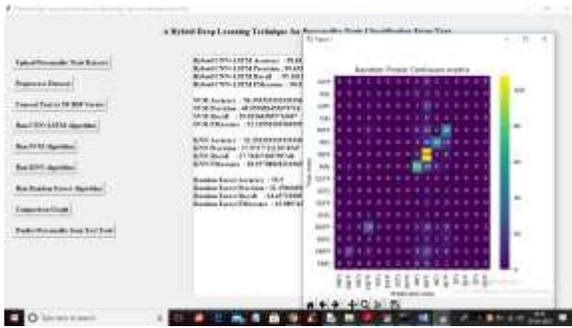
In above screen we can see with CNN+LSTM we got 98% accuracy and in above confusion matrix graph x-axis represents predicted TRAIT classes and y-axis represents original test classes and in diagonal we can see all boxes prediction are correctly matching with original test classes in y-axis and now close above graph and then click on 'Run SVM Algorithm' button to get below output



In above screen with SVM we got 56% accuracy and in confusion matrix graph its showing all wrong prediction and now close above graph and then click on 'Run KNN Algorithm' button to get below output



In above screen with KNN we got 31% accuracy and now click on "Run Random Forest" button to get below output



In above screen first we are displaying user POST message and then after arrow symbol ==> showing classification TRAIT



Similarly by using any other user POST we can predict or classify TRAIT

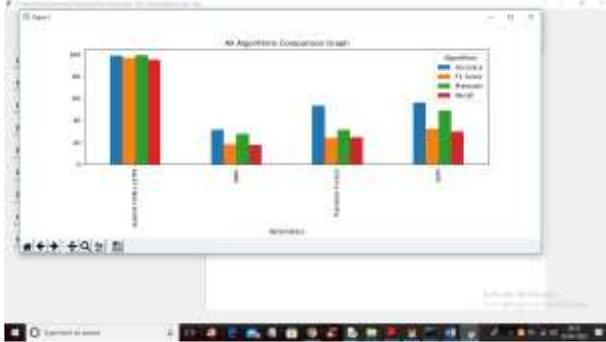
VI. CONCLUSION

Personality trait classification from textual data has become an important research area in cognitive computing, natural language processing, and social media analytics. With the rapid growth of online communication platforms, large volumes of textual content are continuously generated by users. Analyzing this data can provide valuable insights into individual behavior, opinions, and personality characteristics. However, accurately identifying personality traits from textual data remains challenging due to the complexity and variability of natural language.

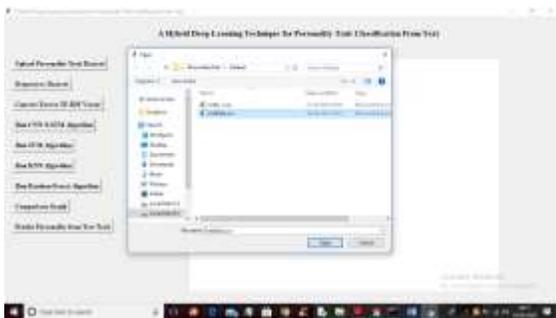
In this work, a Hybrid Deep Learning Approach for Personality Trait Classification from Textual Data has been presented. The proposed system combines Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to effectively capture both local textual features and long-term contextual information. The CNN model is responsible for extracting important semantic patterns from the text, while the LSTM model preserves contextual relationships between words and sentences. By integrating these two models, the hybrid architecture improves the overall performance of personality trait classification.

The experimental results demonstrate that the hybrid CNN-LSTM model achieves better accuracy and classification performance compared to traditional machine learning algorithms such as SVM, KNN, and Random Forest. The system effectively classifies personality traits such as Introversion-Extroversion,

In above screen with Random Forest we got 53% accuracy and now click on 'Comparison Graph' button to get below output



In above graph x-axis represents algorithms names and y-axis represents accuracy and other metrics where different colour bar represents different metrics and in all algorithm propose Hybrid CNN+LSTM got high accuracy and now close above graph and then click on 'Predict Personality from Test Trait' button to upload test data and then Hybrid CNN+LSTM will predict TRAIT from test data



In above screen selecting and uploading 'testData.csv' and then click on 'Open' button to get below output



Intuition–Sensing, Thinking–Feeling, and Judging–Perceiving from textual content. This improved performance highlights the effectiveness of deep learning-based hybrid models for analyzing complex textual data.

Overall, the proposed system provides an efficient and scalable solution for automatic personality detection from textual data. The results of this research can be applied in several real-world applications such as recruitment analysis, customer behavior understanding, recommendation systems, social media analytics, and psychological research. Future work may focus on improving the model by incorporating larger datasets, multilingual text analysis, and advanced transformer-based deep learning architectures for even more accurate personality prediction.

REFERENCE

1. S. Ahmad, M. Z. Asghar, F. M. Alotaibi, and S. Khan, "Classification of poetry text into the emotional states using deep learning technique," *IEEE Access*, vol. 8, pp. 73865–73878, 2020.
2. N. Majumder, S. Poria, A. Gelbukh, and E. Cambria, "Deep learning-based document modeling for personality detection from text," *IEEE Intell. Syst.*, vol. 32, no. 2, pp. 74–79, Mar. 2017.
3. W. A. H. W. Azizan, A. A. A. Rahim, S. L. M. Hassan, I. S. A. Halim, and N. E. Abdullah, "A comparative study of two machine learning algorithms for heart disease prediction system," in *Proc. IEEE 12th Control Syst. Graduate Res. Colloq. (ICSGRC)*, Shah Alam, Malaysia, Aug. 2021, pp. 132–137, doi: 10.1109/ICSGRC53186.2021.9515250.
4. A. Hassan and A. Mahmood, "Convolutional recurrent deep learning model for sentence classification," *IEEE Access*, vol. 6, pp. 13949–13957, 2018.
5. N. Chen and P. Wang, "Advanced combined LSTM-CNN model for Twitter sentiment analysis," in *Proc. 5th IEEE Int. Conf. Cloud Comput. Intell. Syst. (CCIS)*, Nov. 2018, pp. 684–687.
6. A. S. Khan, H. Ahmad, M. Zubair, F. Khan, A. Arif, and H. Ali, "Personality classification from online text using machine learning approach," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 3, pp. 1–5, 2020, doi: 10.14569/IJACSA.2020.0110358.
7. L. Liu, D. Preotiu-Pietro, Z. R. Samani, and M. E. Ungar, "Analyzing personality through social media profile picture choice," in *Proc. Int. AAAI Conf. Social Media (ICWSM)*, 2016, pp. 211–220.
8. All things Statista. Number of Monthly Active Twitter Users Worldwide From 1st Quarter 2010 to 1st Quarter 2019 Retrieved From. Accessed: Jan. 18, 2020. [Online]. Available: <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>
9. N. Majumder, S. Poria, A. Gelbukh, and E. Cambria, "Deep learning-based document modeling for personality detection from text," *IEEE Intell. Syst.*, vol. 32, no. 2, pp. 74–79, Mar. 2017.
10. D. Xue, L. Wu, Z. Hong, S. Guo, L. Gao, Z. Wu, X. Zhong, and J. Sun, "Deep learning-based personality recognition from text posts of online social networks," *Int. J. Speech Technol.*, vol. 48, no. 11, pp. 4232–4246, Nov. 2018.
11. M. Osama and S. R. El-Beltagy, "A transfer learning approach for emotion intensity prediction in microblog text," in *Proc. Int. Conf. Adv. Intell. Syst. Inform. Cham, Switzerland: Springer*, Oct. 2019 pp. 512–522.
12. A. Khattak, M. Z. Asghar, Z. Ishaq, W. H. Bangyal, and I. A. Hameed, "Enhanced concept-level sentiment analysis system with expanded ontological relations for efficient classification of user reviews," *Egyptian Informat. J.*, vol. 3, pp. 1–17, Apr. 2021, doi: 10.1016/j.eij.2021.03.001.
13. M. Z. Asghar, A. Sattar, A. Khan, A. Ali, F. Masud Kundi, and S. Ahmad, "Creating sentiment lexicon for sentiment analysis in Urdu: The case of a resource-poor language," *Expert Syst.*, vol. 36, no. 3, Jun. 2019, Art. no. e12397.
14. F. M. Alotaibi, M. Z. Asghar, and S. Ahmad, "A hybrid CNN-LSTM model for psychopathic class detection from tweeter users," *Cognit. Comput.*, vol. 13, no. 3, pp. 709–723, Mar. 2021.
15. A. Khattak, A. Habib, M. Z. Asghar, F. Subhan, and I. A. R. Habib, "Applying deep neural



networks for user intention identification,” *Soft Comput.*, vol. 25, no. 3, pp. 2191–2220, Feb. 2021.