iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.79311

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

SEBOT-GCL: Graph Contrastive Learning for Social Bot Detection

P. Sujitha[1], J. Yogasri[2], V. Vysyaa[3], S. Venkata Lakshmi[4]
[1, 2, 3]*UG Scholar,* [4]*Associate Professor, Computer Science and Engineering, K.L.N. College of Engineering*

Abstract: ***Social media platforms have experienced a rapid increase in automated accounts known as social bots, which are capable of spreading misinformation, spam, and malicious content. Detecting these bots is essential to maintain the reliability and security of online social networks. This paper proposes an intelligent social bot detection framework that utilizes graph-based learning and contrastive learning techniques to accurately identify automated accounts. The system constructs a social interaction graph where each node represents a user and edges represent interactions between users. A contrastive learning mechanism is used to learn meaningful representations of user behavior from multiple graph views. The proposed model analyzes structural and behavioral patterns to distinguish genuine users from bot accounts. Experimental evaluation performed on benchmark social media datasets demonstrates high accuracy and improved detection performance compared to traditional machine learning models. A user-friendly interface is also developed to visualize datasets, graph structures, and prediction results, allowing researchers and administrators to monitor suspicious activities effectively. The proposed framework provides a reliable and scalable solution for detecting social bots in large-scale social media networks.***

Keywords: Social Bot Detection, Graph Neural Networks, Contrastive Learning, Social Network Analysis, Machine Learning, Behavioral Pattern Analysis.

I. INTRODUCTION

Online social networks such as Twitter, Facebook, and Instagram have become important platforms for communication, information sharing, and public interaction. However, the rapid growth of these platforms has also led to the emergence of automated accounts known as social bots. These bots are designed to imitate human behavior while performing automated tasks such as spreading spam messages, promoting advertisements, manipulating public opinion, or distributing misinformation. The presence of social bots can significantly affect the credibility and reliability of information shared on social media platforms.

Traditional bot detection approaches rely on manual analysis or rule-based systems that examine user activity patterns such as posting frequency, follower counts, and profile attributes. Although these methods can identify simple bots, they often fail to detect sophisticated bots that mimic human-like behavior and interact with other users in complex ways. Additionally, manual detection methods are time-consuming and not suitable for analyzing large-scale social network data.

Recent advances in artificial intelligence and machine learning have enabled the development of automated systems capable of analyzing large volumes of social media data. In particular, graph-based learning methods have shown great potential in analyzing social networks because they capture relationships and interactions between users. Graph Neural Networks (GNNs) can analyze both individual user features and network connectivity patterns, enabling more accurate identification of suspicious accounts.

This project focuses on developing an automated social bot detection system using graph-based contrastive learning techniques. The proposed framework constructs a social interaction graph and learns meaningful user representations by comparing multiple graph views. These learned representations help the system differentiate between genuine users and automated bots based on behavioral and structural patterns. The goal of this research is to provide a reliable and scalable solution for detecting social bots and improving the integrity of social media platforms.

II. METHODOLOGY

The proposed social bot detection system is implemented as an automated machine learning pipeline developed using Python. The system integrates data preprocessing, graph construction, contrastive learning, model training, and classification modules. Python libraries such as Pandas and NumPy are used for dataset handling and numerical operations, while NetworkX is used for graph construction and analysis. Machine learning and graph learning models are implemented using frameworks such as PyTorch and Scikit-learn. Visualization libraries such as Matplotlib and Seaborn are used to generate graphical representations of results.

The methodology begins with data collection from publicly available social media datasets containing user profile attributes and interaction patterns. These datasets include features such as retweet count, mention count, follower relationships, and verification status. The collected data is preprocessed to remove missing values and normalize numerical attributes.

After preprocessing, the system constructs a social interaction graph where nodes represent users and edges represent interactions between them. This graph representation enables the system to capture structural relationships within the social network. A contrastive learning framework is then applied to generate multiple augmented views of the graph. These views allow the model to learn meaningful embeddings that capture both structural and behavioral characteristics of users.

The learned embeddings are used to train a classification model that predicts whether an account is a bot or a genuine user. The model performance is evaluated using standard evaluation metrics such as accuracy, precision, recall, and F1-score. The final system provides prediction results and visualization outputs through a graphical interface, enabling users to analyze the detection results efficiently.

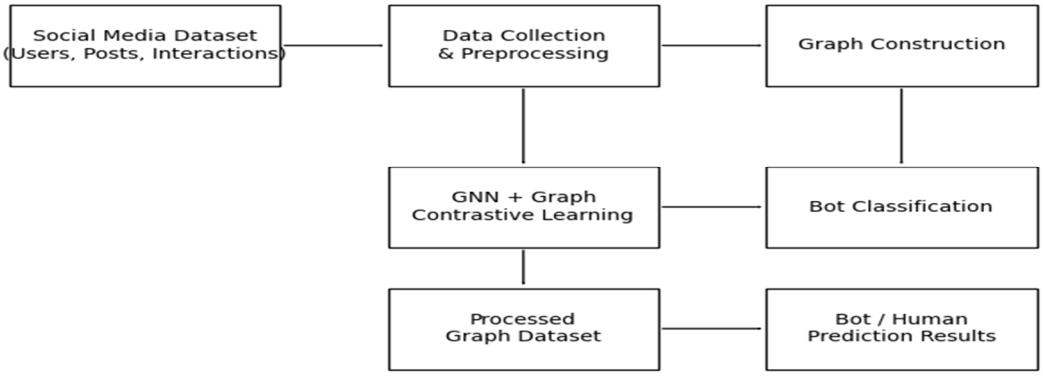


Fig.1 Flow Diagram

III. PREPROCESSING

Data preprocessing is an essential step in preparing social network datasets for machine learning analysis. Raw social media data often contains missing values, inconsistent formatting, and redundant information. These issues can negatively affect the performance of the detection model if not properly addressed.

In this project, preprocessing begins with cleaning the dataset by removing duplicate records and handling missing values. Numerical features such as retweet count, mention count, and follower count are normalized to ensure consistent scaling across the dataset. Categorical attributes such as verification status are converted into numerical representations for model compatibility.

Feature selection is also performed to identify the most relevant attributes for bot detection. Behavioral features such as posting frequency, interaction patterns, and follower relationships provide valuable information for distinguishing between human users and automated accounts. These selected features are then used to construct the graph representation for further analysis.

IV. PROCESS FLOW

The process flow of the proposed social bot detection system begins with data acquisition from social media datasets. The collected data includes user profile information, activity patterns, and interaction relationships between users. These datasets are then passed through a preprocessing stage where data cleaning, normalization, and feature extraction are performed.

Once the data is prepared, the system constructs a social interaction graph representing relationships between users. Each user is represented as a node, while interactions such as mentions, replies, or retweets form edges between nodes. This graph representation captures the connectivity structure of the social network.

The constructed graph is then processed using a contrastive learning framework that generates multiple graph views through augmentation techniques. The model learns meaningful node embeddings by comparing these views and identifying similarities between user behavior patterns. These embeddings are then used by the classification module to determine whether a user account is a bot or a genuine user.

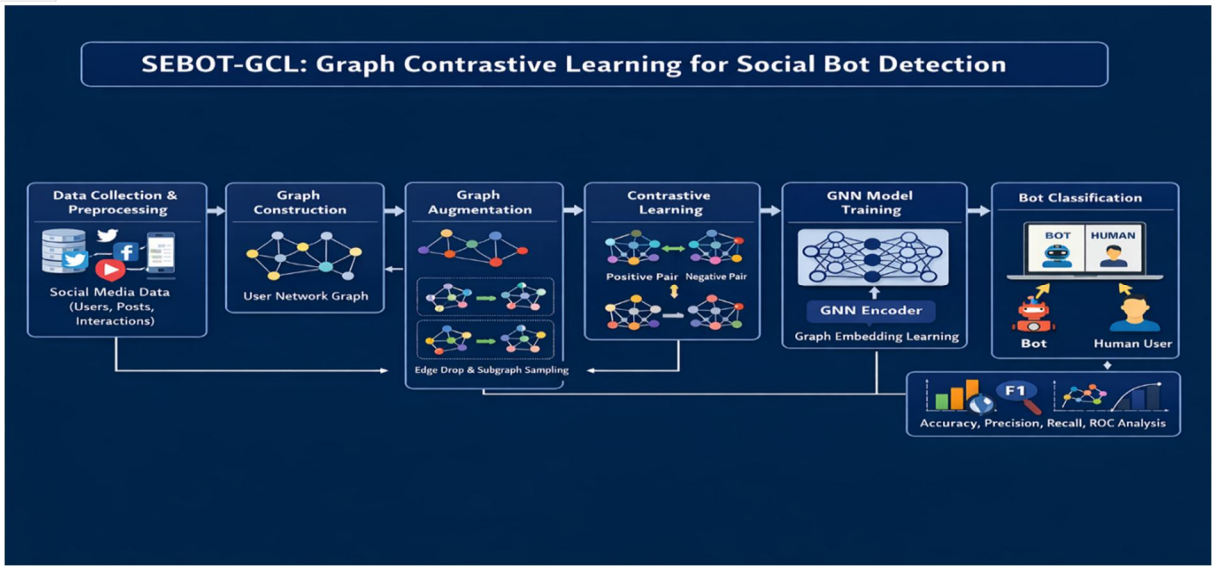


Fig.2 Process Diagram

V. MODEL TRAINING AND AUGMENTATION

To improve model generalization and prevent overfitting, graph augmentation techniques are applied during the training process. These techniques create multiple variations of the social network graph by randomly modifying edges or node attributes while preserving the overall network structure. By training on these augmented graphs, the model learns robust representations of user interactions.

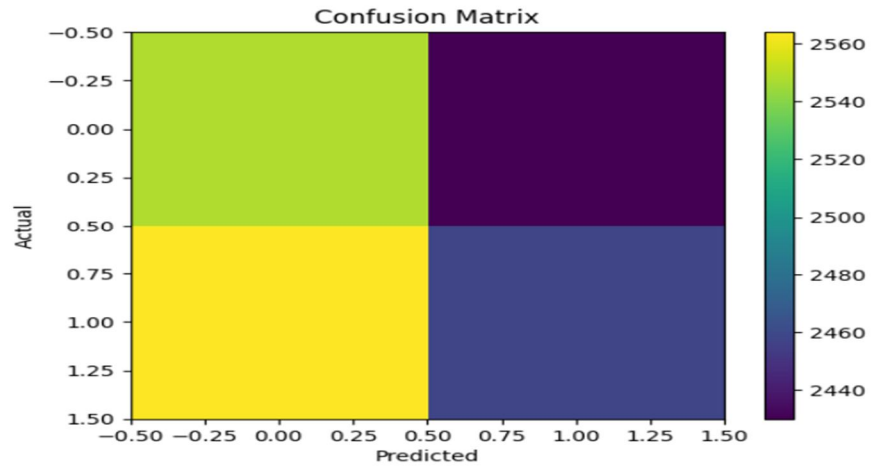
The contrastive learning mechanism compares embeddings generated from different augmented views of the graph. Nodes that represent the same user across different views are treated as positive pairs, while nodes representing different users are treated as negative pairs. This learning strategy helps the model capture meaningful structural patterns within the social network.

The model is trained using supervised learning where labeled data indicates whether an account is a bot or a genuine user. During training, the model optimizes its parameters to minimize classification errors and improve detection accuracy.

```
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt

y_pred = model.predict(X_test)
cm = confusion_matrix(y_test, y_pred)

plt.imshow(cm)
plt.title("Confusion Matrix")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.colorbar()
plt.show()=
```



VI. MATHEMATICAL FOUNDATIONS OF THE PROPOSED METHOD

A. Confusion Matrix

A confusion matrix is used to evaluate the performance of the classification model by comparing predicted labels with actual labels.
Actual / Predicted | Bot | Human
Bot | TP | FN
Human | FP | TN
Where
TP = correctly detected bot accounts
TN = correctly detected genuine users
FP = genuine users misclassified as bots
FN = bot accounts not detected by the model

B. Precision, Recall and F1 Score

Precision measures how many predicted bot accounts are actually bots.
Precision = TP / (TP + FP)
Recall measures how many actual bot accounts are detected.
Recall = TP / (TP + FN)
F1 Score balances precision and recall.
F1 Score = 2 × (Precision × Recall) / (Precision + Recall)

C. Accuracy

Accuracy measures the overall correctness of the classification model.
Accuracy = (TP + TN) / (TP + TN + FP + FN)

VII. RESULT

The experimental results demonstrate that the proposed social bot detection system effectively identifies automated accounts within social networks. The model was trained and evaluated using labeled social media datasets containing both bot accounts and genuine users. The system successfully analyzed behavioral and interaction patterns to classify user accounts with high accuracy.
The confusion matrix generated during evaluation shows a high number of correctly classified accounts, indicating that the model is capable of distinguishing between bot accounts and genuine users. The visualization results also demonstrate clear separation between the two classes within the learned feature space.Overall, the proposed approach improves detection accuracy compared to traditional machine learning methods by incorporating graph-based learning and contrastive learning techniques.

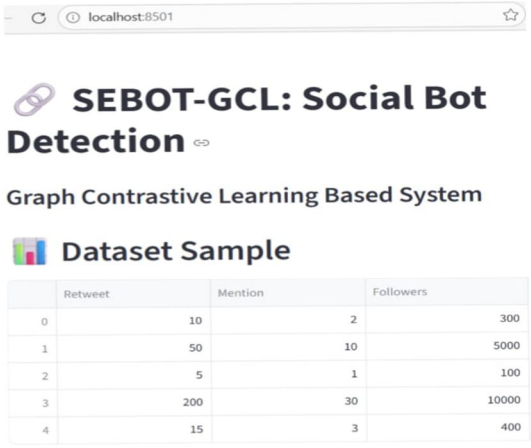


Fig.4 Dataset Sample

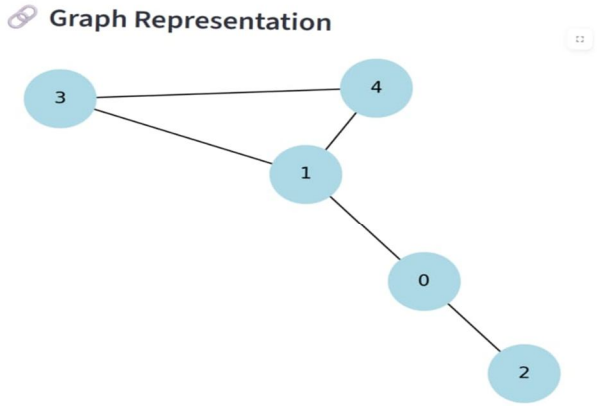


Fig.5 Graph Reprentation

VIII. CONCLUSION

In conclusion, this research successfully developed an automated social bot detection system using graph-based machine learning techniques. The proposed framework integrates data preprocessing, graph construction, contrastive learning, and classification into a unified pipeline capable of detecting automated accounts in social media networks.By leveraging graph representations and contrastive learning strategies, the model captures both structural and behavioral patterns of user interactions, leading to improved detection accuracy. The system demonstrates strong performance when evaluated using standard classification metrics such as accuracy, precision, recall, and F1-score.

The developed framework provides a practical tool for researchers and social media administrators to monitor suspicious activities and maintain the integrity of online platforms. Future work may focus on extending the system to support real-time monitoring and cross-platform analysis for enhanced social media security.

REFERENCES

[1] W. E. Zhang, Q. Z. Sheng, A. Alhazmi, and C. Li, ‘‘Adversarial attacks on deep-learning models in natural language processing: A survey,’’ ACM Trans. Intell. Syst. Technol. (TIST), vol. 11, no. 3, pp. 1–41, 2020.

[2] D. I. Adelani, H. Mai, F. Fang, H. H. Nguyen, J. Yamagishi, and I. Echizen, ‘‘Generating sentiment-preserving fake online reviews using neural language models and their human-and machine-based detection,’’ in Proc. 34th Int. Conf. Adv. Inf. Netw. Appl., 2020, pp. 1341–1354.

[3] M. Aljabri, R. Zagrouba, A. Shaahid, F. Alnasser, A. Saleh, and D. M. Alomari, Machine Learning—Based Social Media Bot Detection: A Comprehensive Literature Review. Cham, Switzerland: Springer, 2023.

[4] Ali and A. M. Syed, ‘‘Cyberbullying detection using machine learning,’’ Pakistan J. Eng. Technol., vol. 3, no. 2, pp. 45–50, Apr. 2022.

[5] J.Wise. Twitter Bots Percentage: HowMany Botsare on Twitter? Earth Web. Accessed: Jun. 14, 2024. [Online]. Available: https://earthweb.com/blog/how-many-bots-are-on-twitter/

[6] W. Yue and L. Li, ‘‘Sentiment analysis using Word2vec-CNN-BiLSTM classification,’’ in Proc. 7th Int. Conf. Social Netw. Anal., Manage. Secur. (SNAMS), Dec. 2020, pp. 1–5.

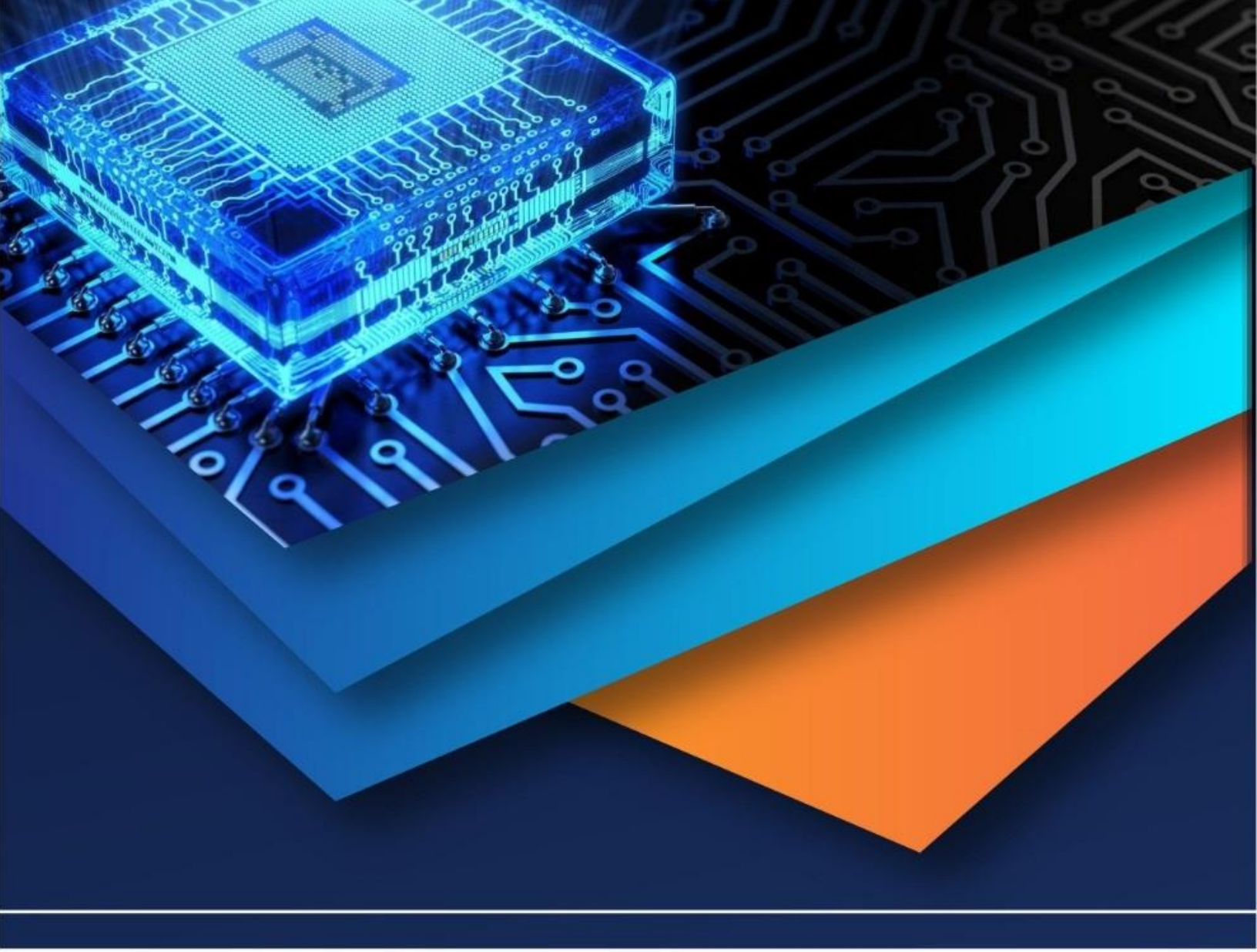
[7] J. Pennington, R. Socher, and C. Manning, ‘‘Glove: Global vectors for word representation,’’ in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2014, pp. 1532–1543.

[8] S. S. Roy, A. I. Awad, L. A. Amare, M. T. Erkihun, and M. Anas, ‘‘Multimodel phishing URL detection using LSTM, bidirectional LSTM, and GRU models,’’ Future Internet, vol. 14, no. 11, p. 340, Nov. 2022.

[9] T. B. Brown et al., ‘‘Language models are few-shot learners,’’ in Proc. NIPS, 2020, pp. 1877–1901.

[10] M. Tezgider, B. Yildiz, and G. Aydin, ‘‘Text classification using improved bidirectional transformer,’’ Concurrency Comput., Pract. Exper., vol. 34, no. 9, p. 6486, Apr. 2022.

[11] S. Kumar and A. Solanki, ‘‘An abstractive text summarization technique using transformer model with self-attention mechanism,’’ Neural Comput. Appl., vol. 35, no. 25, pp. 18603–18622, Sep. 2023.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY