ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

Robust Malicious Query Access Detection using GCNN-Attention and BiLSTM with DFA Analysis

Ms Donepudi Prasanthi[1], Dr.Juluru Pavan Sai[2], Sevugarajan.S[3], Shaik Ayesha[4], Sateesh Yalavarthi[5]

#1 Post Graduate Scholar, Department of CSE, Vikas College of Engineering & Technology, Nunna, Andhra Pradesh, India- 521212.

#2 Associate Professor, Department of CSE, Vikas College of Engineering & Technology, Nunna, Andhra Pradesh, India- 521212.

#3 Professor, Department of ECE, Vikas College of Engineering & Technology, Nunna, Andhra Pradesh, India- 521212.

#4 Post Graduate Scholar, Department of CSE, Vikas College of Engineering & Technology, Nunna, Andhra Pradesh, India- 521212.

#5 Post Graduate Scholar, Department of CSE, Vikas College of Engineering & Technology, Nunna, Andhra Pradesh, India- 521212.

Mail Id: donepudi.prasanthi@gmail.com [1], julurupavansai@gmail.com [2], dean_academics@vikasinstitutionsnunna.org[3] , ayeshamca333@gmail.com [4], sateeshthermal@gmail.com [5]

Abstract: Language technology resource repositories are increasingly vulnerable to illicit query access in the current digital environment, which poses major hazards to intellectual property and data security. In order to efficiently identify unusual query behavior, this study suggests a hybrid deep learning architecture that combines Graph Convolutional Neural Network (GCNN) with Attention mechanism and Bidirectional Long Short-Term Memory (BiLSTM). TF-IDF feature extraction is used to alter textual queries, allowing for effective representation for categorization. Detrended Fluctuation Analysis (DFA) is used to differentiate between benign and malevolent access patterns in order to further capture long-range behavioral patterns.

According to experimental data, the suggested model outperforms conventional machine learning techniques like SVM and Naïve Bayes, with an

accuracy of 93.85%. A reliable and scalable method for protecting language technology resources from changing cyberthreats is the combination of deep learning and long-range correlation analysis.

Index terms *- — Malicious Query Detection, Language Technology Resources, Deep Learning, Graph Convolutional Neural Network (GCNN), Attention Mechanism, Bidirectional Long Short-Term Memory (BiLSTM), TF-IDF, Detrended Fluctuation Analysis (DFA), Cybersecurity, Access Behavior Analysis*

1. INTRODUCTION

In contemporary digital systems, language technology resources are essential for applications like text analytics, intelligent learning platforms, and information retrieval. These repositories confront growing security issues as they become more publicly accessible in academic, commercial, and public sectors. In particular, the rise of automated and malicious query access threatens data confidentiality, intellectual property, and overall system reliability.

Conventional security measures rely on machine learning methods that examine isolated query characteristics, including Support Vector Machines and Naïve Bayes. Nevertheless, these techniques are less successful against sophisticated assaults because they are unable to identify intricate language patterns and long-range behavioral connections in user access sequences. Accurately recognizing malicious query activity is severely hampered by the incapacity to model contextual and sequential linkages.

This research suggests a hybrid deep learning architecture that combines Graph Convolutional Neural Network (GCNN) with an Attention mechanism and Bidirectional Long Short-Term Memory (BiLSTM) to overcome these drawbacks. The system can capture the temporal, contextual, and structural characteristics of query data thanks to this combination. To further improve detection reliability, long-range correlations in user behavior are analyzed using Detrended Fluctuation Analysis (DFA).

The goal of the suggested method is to offer a reliable, scalable, and clever way to identify unauthorized query access in language technology repositories. The solution helps to boost cybersecurity and provide safe access to important language resources by increasing classification accuracy and comprehending user behavior patterns.

2. LITERATURE SURVEY

a) AI-Driven Safety and Security for UAVs: From Machine Learning to Large Language Models:

As the applications of unmanned aerial vehicles (UAVs) expand beyond emergency response, agriculture, and logistics, safety and security threats become more complex. To address these evolving risks, which include physical safety and network security concerns, it is vital to continue developing by adding traditional artificial intelligence (AI) methods like machine learning (ML) and deep learning (DL), which significantly increase UAV safety and security. Large language models (LLMs), a cutting-edge advancement in artificial intelligence, are associated with strong learning and adaptability in a range of scenarios. Their emergence is a part of a broader trend toward intelligent systems that may eventually display human-like thought processes. This paper discusses the evolution of traditional AI technologies as documented in the literature, identifies typical security and safety threats that affect UAVs, and offers solutions to mitigate such threats. It

also highlights the drawbacks of traditional AI technology and examines the current status of LLM use in UAV safety and security. This work concludes by discussing the challenges and possible future research directions for improving UAV safety and security using LLMs. By leveraging their state-of-the-art capabilities, LLMs offer potential benefits in critical domains such as emergency response, precision agriculture, and urban air traffic management, fostering revolutionary developments toward safe, dependable, and flexible UAV systems that address modern operational challenges.

b) Signature-based intrusion detection using machine learning and deep learning approaches empowered with fuzzy clustering:

Network security is crucial in the current digital era since sensitive data and vital infrastructure are always at danger. By combining deep learning (DL) and machine learning (ML) methods for instruction detection, this research seeks to improve network security. By breaking into networks and gathering personal information, attackers have tried to undermine security measures.Intrusion detection systems (IDSs), which monitor and analyze data with the aim of identifying and reporting unsafe activities to help stop an attack, are a crucial part of cybersecurity.Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), Decision Tree (DT), Long Short-Term Memory (LSTM), and Artificial Neural Network (ANN) are some of the vector figures used in the study. These models are tested in order to get the best results for detecting and stopping network infractions. The results gathered show that all of the models under investigation are capable of organizing data derived from network traffic. In order to distinguish between normal and intrusive activities, models such as SVM, KNN, RF, and DT have demonstrated good results. Deep learning models LSTM and ANN rapidly detect complex and long-term patterns in network data. It is very helpful for managing complex invasions due to its exceptional precision, accuracy, and memory.Our research indicates that the flexibility and explainability of SVM and Random Forest make them good choices for real-world IDS applications. Businesses searching for trustworthy and simpler-to-understand IDS solutions may find these models helpful. Furthermore, because LSTM and ANN can record consecutive occurrences, they are suitable for scenarios with complex, developing threats.

c) Comparative analysis of deep learning and traditional methods for IoT botnet detection using a multi-model framework across diverse datasets:

The expansion of Internet of Things (IoT) devices has created previously unheard-of cybersecurity vulnerabilities, making botnets a major threat to network infrastructure. This study uses a weighted soft-voting technique to merge Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (BiLSTM), Random Forest (RF), and Logistic Regression (LR) to address these issues. It emphasizes both deep learning and traditional machine learning methods. Our approach offers a Quantile Uniform transformation to reduce feature skewness, an individual deep learning–traditional machine learning performance, a hybrid model (ensemble models) for robust detection, and a multi-layered feature selection methodology to increase discriminative strength. The system achieves 100% accuracy on BOT-IOT, 99.2% on CICIOT2023, and 91.5% on IOT23, outperforming state-of-the-art models by up to 6.2% on these datasets. These contributions increase IoT security by offering

scalable, high-performance detection that is adaptable to various network situations and provides practical enhancements for real-world application.

d) Enhancing Multilingual Hate Speech Detection: From Language-Specific Insights to Cross-Linguistic Integration:

Hate speech against sexual orientation, gender, race, and religion can be spread on social media by prejudiced individuals. While unfavorable remarks might harm one's social standing and mental health, constructive contacts in various civilizations can increase self-esteem. To lessen hazardous content's negative effects on individuals and communities, it is crucial to identify and remove it. The rising frequency highlights the necessity for stricter laws and policies pertaining to digital platforms in order to protect people from such dangerous conduct. The goal of hate speech is typically to denigrate or alienate a group because of their identity. Research on hate speech concentrates on resource-aware languages, such as Chinese, German, and English. Languages with limited resources, such as Italian, Spanish, Portuguese, Roman Urdu, Korean, and Indonesian, provide difficulties. Information extraction is more difficult when there are insufficient linguistic resources. The goal of this project is to identify and improve hate speech in thirteen different languages. To finish our analysis, we examined transformer-based algorithms, mainstream deep learning, and conventional machine learning. Strong and universal hate speech recognition across dialects was achieved by hyperparameter adjustment, optimization, and generative configurations. We enhanced detection performance by up to 10% in both accuracy and recall in several understudied languages. By offering deeper insights into model choices, we also enhance explainable AI in this setting, which is crucial for the ethical and regulatory implementation of AI. Our work demonstrates notable performance gains across datasets and languages through meticulous comparisons. With F1-scores of 0.90 in German (GermEval-2018), up from 0.72, and 0.93 in German (GermEval-2021), up from 0.58, our model outperformed benchmarks. Additionally, it improved from 0.91 to 0.95 in Roman Urdu High School. Our accuracy improved from 0.59 to 0.96 for mixed-language datasets such as Italian and English (AMI 2018). The robustness and flexibility of our model established a new standard for hate speech detection systems in various linguistic situations.

e) Effective Detection of Malicious Uniform Resource Locator (URLs) Using Deep-Learning Techniques:

Because of the rapid increase in internet usage, malicious URLs are a common cybercrime. Traditional detection techniques typically have high false alarm rates and are unable to keep up with emerging threats because of outdated feature extraction techniques and datasets. To get over these limitations, we introduce a deep learning-based approach for detecting fraudulent URLs. In the Char2B model, a Conv1D layer with a three-kernel kernel, unit-sized stride, and padding is combined with BERT and CharBiGRU embedding. We compared to the BERT model after embedding. PhishTank, Any, and the open project directory (DMOZ).For the study, Run supplied 87,216 malicious and benign URLs. Accuracy, precision, recall, and F1-score were used to test the models after they were trained on the training set. Model efficacy and performance were optimized by iterative

refinement. With 98.50% accuracy, 98.27% precision, 98.69% recall, and 98.48% F1-score, our model performed better than the baseline BERT model. The baseline had 0.018 false positives, while our model had 0.017. In order to increase detection, the system extracted and used helpful information to identify URLs as either benign or harmful. This study demonstrates how our deep learning approach enhances cybersecurity by integrating advanced algorithms that boost digital safety, defense mechanisms, and detection accuracy.

3. METHODOLOGY

i) Proposed Work:

By combining machine learning, deep learning, and behavioral analysis approaches, the proposed study presents a strong and intelligent framework for identifying fraudulent query access in language technology resource repositories. In order to increase data quality, the system first gathers query datasets from sources like Kaggle. Next, preparation procedures like stopword and special character removal and normalization are carried out. TF-IDF vectorization is then used to convert cleaned textual input into numerical form, allowing for efficient feature representation for model training.

The framework uses both conventional and sophisticated models to carry out categorization. While a deep learning architecture that combines Graph Convolutional Neural Network (GCNN) with an Attention mechanism is used to capture complex relationships and feature importance within query data, machine learning algorithms like Support Vector Machine (SVM) and Naïve Bayes are used as baseline models. A Bidirectional Long Short-Term Memory (BiLSTM) layer is added to further improve the comprehension of user activity patterns by capturing sequential dependencies in both forward and backward directions.

Furthermore, the system uses Detrended Fluctuation Analysis (DFA) to examine long-range correlations in query access behavior, allowing behavioral trends to be used to distinguish between fraudulent and legitimate activity. Real-time query prediction is supported via a Flask-based web interface that enables users to upload data and receive instant categorization results. In order to secure language technology resources, this hybrid strategy guarantees increased detection accuracy, scalability, and useful deployment.

ii) System Architecture:

To identify unauthorized query access in language technology resources, the system design employs a structured pipeline. User queries are first supplied as input and go through preprocessing, which eliminates noise like stopwords and unusual letters. TF-IDF is then used to transform the cleaned data into numerical feature vectors, allowing textual information to be effectively represented. The core detection module, which combines a Graph Convolutional Neural Network (GCNN) with an Attention mechanism and a BiLSTM layer, receives these characteristics. In order to accurately classify queries, this hybrid model incorporates structural linkages, significant contextual factors, and sequential dependencies within query data.

After the classification phase, the system uses Detrended Fluctuation Analysis (DFA) for long-range correlation analysis to look at trends in user behavior over time. The system gives evaluation criteria for performance assessment and classifies

outputs into malicious or valid queries based on the learnt patterns. As the above architectural diagram illustrates, the combination of deep learning and behavioral analysis guarantees a reliable, scalable, and real-time detection system that can protect linguistic resource repositories from complex threats.

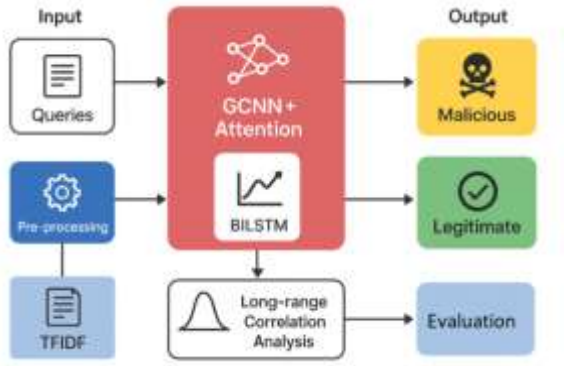

Fig1 proposed architecture

iii) Modules:

1. Packages Imported:

By importing the necessary Python libraries for data processing, model creation, and visualization, this module initializes the system. It guarantees a steady environment for effectively carrying out deep learning and machine learning tasks.

2. Examining the Dataset

To comprehend query distribution, class labels, and data features, the dataset is examined. In order to inform preprocessing and model design choices, this stage assists in identifying imbalance, noise, and trends.

3. Visualization

Plots and charts are examples of graphic representations that are used to track class distribution and query behavior. Visualization facilitates improved model tuning and enhances comprehension of data patterns.

4. Vectorization of TF-IDF:

TF-IDF is used to transform textual queries into numerical feature vectors. Linguistic data may be efficiently processed by machine learning and deep learning models because to this transformation.

5. Train and Test Data Splitting:

To assess the model's performance on untested data, the dataset is split into training and testing sets. This guarantees that the model does not overfit and has good generalization.

6. Training Models:

Extracted features are used to train several models, including SVM, Naïve Bayes, and the suggested GCNN + Attention + BiLSTM. This module makes it possible to discover patterns that differentiate between legal and fraudulent requests.

7. Assessment

Metrics like accuracy, precision, recall, and F1-score are used to assess the trained models. This aids in performance comparison and model selection.

8. The Flask Server

Flask is used to provide a web-based interface that enables real-time communication. Users may submit inquiries and get predictions right away.

9. Log in as a user:

To guarantee that only those with permission may access the system, authentication is used. This improves security and safeguards private information.

10. Language Query Access Behavior Identification:

In order to categorize input queries as dangerous or genuine, this fundamental module uses the learned

model and DFA analysis. It offers behavior insights in addition to final forecasts.

11. Log off:

By safely ending user sessions, this module preserves overall system security and stops unwanted access after system usage.

iv) ALGORITHMS:

1. Support Vector Machine (SVM):

As a baseline supervised learning approach, Support Vector Machine (SVM) is used to categorize requests into harmful and lawful groups. It finds an ideal hyperplane that optimizes the margin between classes by transferring TF-IDF feature vectors into a high-dimensional space. This method guarantees tolerance against noise in textual data and good generalization capabilities. SVM offers a trustworthy benchmark for assessing the performance of sophisticated deep learning models because of its efficiency in managing sparse data.

2. Naïve Bayes:

Based on Bayes' theorem, Naïve Bayes is a probabilistic classification technique that assumes feature independence. By taking into account the probability of feature occurrences in the dataset, it determines the posterior probability of each class. Because it can handle vast vocabularies with little complexity, this approach is computationally efficient and especially well-suited for text categorization applications. It offers competitive performance and is a helpful reference for more complicated devices despite its simplicity.

3. GCNN + Attention:

By capturing both global and local interactions inside textual queries, the Graph Convolutional Neural Network (GCNN) in conjunction with an Attention mechanism improves feature extraction. While the Attention method gives the most pertinent characteristics larger weights, GCNN models the structural relationships between words, enabling the model to concentrate on significant portions of the input. By efficiently identifying contextual patterns and lessening the influence of irrelevant data, this combination increases classification accuracy.

4. GCNN + Attention + BiLSTM (Proposed Model):

The suggested hybrid model achieves better performance in identifying fraudulent queries by combining GCNN, Attention, and Bidirectional Long Short-Term Memory (BiLSTM). The Attention method emphasizes important components, GCNN extracts structural characteristics, and BiLSTM records sequential dependencies in both forward and backward directions. A better comprehension of query behavior is made possible by this multi-layered architecture, which also improves overall accuracy and dependability and makes it easier to identify subtle and intricate harmful patterns.

5. Detrended Fluctuation Analysis (DFA):

Long-range correlations in user query access behavior over time are examined using Detrended Fluctuation Analysis (DFA). To differentiate between typical and unusual activity, it calculates scaling parameters and assesses changes in query patterns. Higher correlation levels are usually associated with malevolent activity, whilst lower correlation values indicate appropriate use. The approach improves behavioral analysis and fortifies the entire detection framework by combining DFA with deep learning outputs.

4. EXPERIMENTAL RESULTS

A query dataset comprising both harmful and genuine samples was used to assess the suggested system. Standard assessment criteria including accuracy, precision, recall, and F1-score were used in performance analysis. To guarantee objective assessment, the dataset was split into training and testing sets. Traditional machine learning models (SVM and Naïve Bayes) and the suggested deep learning models (GCNN + Attention and GCNN +

Attention + BiLSTM) were compared in experiments. The findings show that because deep learning techniques can identify sequential and contextual patterns in query data, they perform noticeably better than conventional techniques.

With an accuracy of 93.85%, precision of 93.04%, recall of 94.19%, and F1-score of 93.53%, the suggested hybrid model (GCNN + Attention + BiLSTM) demonstrated exceptional detecting capacity. Sequence learning was strengthened by the addition of BiLSTM, and harmful query patterns were more easily identified thanks to DFA's improved behavioral analysis. In comparison to current methods, the suggested system offers more accuracy, fewer false positives, and enhanced resilience, as demonstrated by graphical comparisons and performance tables (included in the document's findings section).

Accuracy: A test's accuracy is determined by its capacity to distinguish between healthy and ill cases. To gauge the accuracy of the test, find the percentage of examined instances that had true positives and true negatives. According to the computations:

Accuracy = TP + TN /(TP + TN + FP + FN)

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: Precision is the number of affirmative cases or the classification's accuracy rate. The following formula is applied to assess accuracy:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$

Recall: A model's ability to recognise every instance of a pertinent machine learning class is measured by its recall. The ratio of accurately predicted positive observations to the total number of positives indicates how well a model can identify class instances.

$$Recall = \frac{TP}{(FN + TP)}$$

mAP: Mean Average Precision is one ranking quality metric (MAP). It considers the number of relevant recommendations and their position on the list. MAP at K is calculated as the arithmetic mean of the Average Precision (AP) at K for each user or query.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$AP_k = \text{the AP of class } k$$

$$n = \text{the number of classes}$$

F1-Score: An accurate machine learning model is indicated by a high F1 score. combining precision and recall to increase model correctness. The accuracy statistic quantifies the frequency with which a model correctly predicts a dataset.

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$

Fig2 upload input

Fig3 results

5. CONCLUSION

This research uses a hybrid deep learning technique to propose a strong and intelligent framework for identifying unauthorized query access in language technology resource repositories. The method greatly increases detection accuracy by efficiently capturing structural, contextual, and sequential patterns in query data with the integration of GCNN, Attention mechanism, and BiLSTM. By examining long-range user behavior, Detrended Fluctuation Analysis (DFA) improves the system even more and makes it possible to distinguish between harmful and legal activity.

According to experimental data, the suggested model achieves excellent accuracy and superior overall performance compared to conventional machine learning techniques. Deep learning and behavioral analysis work together to secure language technology resources in a scalable, effective, and useful way. By guaranteeing secure, dependable, and regulated access to sensitive data, our effort strengthens cybersecurity in digital linguistic platforms.

6. FUTURE SCOPE

In order to increase contextual comprehension of complicated query patterns, the suggested system can be further improved by including sophisticated transformer-based models like BERT or GPT. Deeper semantic linkages may be captured by these models, resulting in even greater detection accuracy for complex malicious queries. Real-time streaming data analysis can also facilitate ongoing user behavior monitoring and quicker threat detection.

In order to effectively manage large-scale language repositories, future work may also concentrate on implementing the system in distributed and cloud systems. By offering comprehensible insights into model decisions, explainable AI (XAI) approaches can enhance transparency. The framework's applicability across many platforms and dynamic cybersecurity scenarios will also be improved by expanding it to include multilingual query detection and adaptive learning processes.

REFERENCES

1) Yang, Z., Zhang, Y., Zeng, J., Yang, Y., Jia, Y., Song, H., ... & An, J. (2025). AI-Driven safety and security for UAVs: From machine learning to large language models. Drones, 9(6), 392.
2) Ahmed, U., Nazir, M., Sarwar, A., Ali, T., Aggoune, E. H. M., Shahzad, T., & Khan, M. A. (2025). Signature-based intrusion detection using machine learning and deep learning approaches empowered with fuzzy clustering. Scientific Reports, 15(1), 1726.
3) Ullah, S., Wu, J., Lin, Z., Kamal, M. M., Mostafa, H., Sheraz, M., & Chuah, T. C. (2025). Comparative analysis of deep learning and traditional methods for IoT botnet detection using a multi-model framework across diverse datasets. Scientific Reports, 15(1), 31072.

4) Hashmi, E., Yayilgan, S. Y., Hameed, I. A., Yamin, M. M., Ullah, M., & Abomhara, M. (2024). Enhancing multilingual hate speech detection: From language-specific insights to cross-linguistic integration. IEEE Access.
5) Munaye, Y. Y., Workneh, A. B., Chekol, Y. B., & Mekonen, A. M. (2025). Effective Detection of Malicious Uniform Resource Locator (URLs) Using Deep-Learning Techniques. Algorithms, 18(6), 355.
6) S. B. Naqvi, M. Afzaal, and G. Qiang, "Editorial: Language, corpora, and technology in applied linguistics," Frontiers Psychol., vol. 14, pp. 1–4, Nov. 2023, doi: 10.3389/fpsyg.2023.1325925.
7) L. Wiechetek, "Unmasking the myth of effortless big data-making an open source multilingual infrastructure and building language resources from scratch," in Proc. 13th Int. Conf. Lang. Resour. Eval. (LREC), Marseille, France, 2022, pp. 1167–1177.
8) J. Mariani, "Developing language technologies with the support of language resources and evaluation programs," Lang. Resour. Eval., vol. 39, no. 1, pp. 35–44, Feb. 2005, doi: 10.1007/s10579-005-2694-3.
9) Z. Peng, F. Liang, and L. Mu, "Big data-based access control system in educational information security assurance," Wireless Commun. Mobile Comput., vol. 2022, pp. 1–7, Jan. 2022, doi: 10.1155/2022/2853821.
10) R. Asija and R. Nallusamy, "Security and complexity analysis of user and data based access control (UDBAC) model," in Proc. Int. Conf. Current Trends Comput., Electr., Electron. Commun. (CTCEEC), Mysore, India, Sep. 2017, pp. 358–366, doi: 10.1109/CTCEEC.2017.8455187.
11) [11] S. Alves and M. Fernández, "A graph-based framework for the analysis of access control policies," Theor. Comput. Sci., vol. 685, pp. 3–22, Jul. 2017, doi: 10.1016/j.tcs.2016.10.018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304397516305965
I. Santos, F. Brezo, X. Ugarte-Pedrero, and P. G. Bringas, "Opcode sequences as representation of executables for data-mining-based unknown malware detection," Inf. Sci., vol. 231, pp. 64–82, May 2013.
12) Santos, C. Laorden, and P. G. Bringas, "Collective classification for unknown malware detection," in Proc. Int. Conf. Secur. Cryptogr., Seville, Spain, Jul. 2011, pp. 251–256.
13) W. Zhang, X. Liang, Y. Zhang, and H. Su, "Sensitive data classification of imbalanced short text based on probability distribution BERT in electric power industry," in Proc. Int. Conf. Pattern Recognit., Mach. Vis. Intell. Algorithms (PRMVIA), Mar. 2023, pp.

169–174, doi: 10.1109/PRMVIA58252.2023.00034.

14) H. Pei, J. Jia, W. Guo, B. Li, and D. Song, "TextGuard: Provable defense against backdoor attacks on text classification," in Proc. Netw. Distrib. Syst. Secur. Symp., San Diego, CA, USA, 2024, pp. 1–18.
15) X. Pang, Y. Su, W. Tao, Z. Hu, and H. Chen, "A network security situational awareness system for text classification," J. Phys., Conf. Ser., vol. 2358, no. 1, Oct. 2022, Art. no. 012012, doi: 10.1088/1742-6596/2358/1/012012.
16) M. Khadhraoui, H. Bellaaj, M. B. Ammar, H. Hamam, and M. Jmaiel, "Survey of BERT-base models for scientific text classification: COVID-19 case study," Appl. Sci., vol. 12, no. 6, p. 2891, Mar. 2022.
17) K. Kusakin, O. V. Fedorets, A. Y. Romanov, "Classification of short scientific texts," Sci. Tech.Inf. Proc., vol. 50, pp. 176–183, 2023, doi: 10.3103/S0147688223030024.
18) N. Qun, X. Li, X. Qiu, and X. Huang, "End-to-end neural text classification for Tibetan," in Proc. Int. Symp. Natural Lang. Process. Based Naturally Annotated Big Data, 2017, pp. 472–480.
19) H. Wang, Z. B. Zhao, Y. Y. Li, and X. Q. Zhang, "Construction and application of GCN model for text classification with associated information," Data Anal. Knowl. Discov., vol. 5, no. 9, pp. 31–41, 2021, doi: 10.11925/infotech.2096-3467.2021.0266.
20) Babburi, S. (2024). Explainable AI Framework for Policy-Compliant Anomaly Detection in Data Pipelines.
21) Todupunuri, A. (2025). IMPROVING CUSTOMER EXPERIENCE WITH MODERN BANKING SOLUTIONS. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5120615
22) Gaddam, S. From Fixed Specifications to Self-Adapting Systems: A Machine Learning Perspective on Software Engineering.
23) Vasagam, M. (2024, August 30). Ensuring security in modern data pipelines: Practical strategies for data engineers. International Journal of Intelligent Systems and Applications in Engineering, 12(22s), 2401.
24) Mudusu, S. K., & Gentyala, S. (2026). Zero-Trust Data Pipelines for AI Systems: A Framework for Secure, Verifiable, and Auditable Data Engineering. JOURNAL OF RECENT TRENDS IN COMPUTER SCIENCE AND ENGINEERING (JRTCSE), 14(2), 10-25.
25) Santthosh Saai Reddy Purmani. (2026). Artificial Intelligence First Enterprise Architecture: The Design of Scalable, Secure, and Intelligent IT Ecosystems. American Journal of AI Cyber Computing Management, 6(1(2)), 1–8. https://doi.org/10.64751/ajaccm.2026.v6.n1(2).pp1-8
26) Purmani, S. S. R. (2025). Streamlining IT operations and service management with agile frameworks. European Journal of

Advances in Engineering and Technology, 12(4), 76–81.

27) Purmani, S. S. R. (2025). Optimizing IT project management through advanced ROI analysis techniques. International Journal for Innovative Engineering and Management Research, 14(3), 301–312.
28) Kotte, G. (2025). Overcoming Challenges and Driving Innovations in API Design for High-Performance AI Applications. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283649
29) Purmani, S. S. R. (2024). Aligning IT investment decisions with overall business strategy from an enterprise program management perspective, focusing on the integration of IT leadership in strategic decision-making processes. International Journal of Communication Networks and Information Security, 16(5), 1213–1219
30) Mahtabi, M., Roshan, M., Muhit, M. M. I., Behvar, A., & Haghshenas, M. (2026). Cryogenic ultrasonic fatigue: Mechanisms, advancements, and insights. Cryogenics, 153, 104257. https://doi.org/10.1016/j.cryogenics.2025.104257
31) Purmani, S. S. R. (2025). Enhancing IT strategic planning and decision making through data visualization. International Journal of Enhanced Research in Management & Computer Applications, 14(4), 75–81
32) Mudusu, S. (2025). Health Insurance Fraud Detection: The Role Of Advanced It Systems In Preventing And Identifying Fraud. International Journal, 16(1), 3769-3777
33) GIRISH KOTTE. (2025). ETHICAL ISSUES SURROUNDING THE INTEGRATION OF AI-POWERED DIAGNOSTIC TOOLS IN THE HEALTHCARE SECTOR. American Journal of AI Cyber Computing Management, 5(4), 329–334. https://doi.org/10.64751/ajaccm.2025.v5.n4.pp329-334
34) Manoharan, D. (2026). AI-Driven Anomaly Detection Models for Preventing Claims Denials and Revenue Leakage in Healthcare. Available at SSRN 6385759.
35) Mudusu, S. K. (2025). AI-driven data engineering in the Internet of Things: Scaling data pipelines for smart device ecosystems. ISCSITR-International Journal of Data Engineering (ISCSITR-IJDE), 6(1), 1–9.
36) Maturi, S. Y. (2025). Vulnerabilities in the 802.11 Wireless Client Selection Mechanis.
37) Agrawal, A. M., Gajula, S., Shinde, R. P., Shah, H., & Ghosh, H. (2025, July). Machine Translation for Long Sequences with Enhanced Attention Mechanisms. In 2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET) (pp. 1-6). IEEE.
38) Subramanian, V. K., Bhambri, S., & Gajula, S. (2025, April). Disentangled Graph Variational Auto-encoder Based Framework to Improve the Operational Efficiency in Cloud Computing

Environments. In International Conference on Computer Vision and Robotics (pp. 396-407). Cham: Springer Nature Switzerland.

39) Maturi, S. Y. (2023). Crowdsourced frontier: Unveiling autonomous adversarial cybercapabilities via open AI competition. International Journal of Intelligent Systems and Applications in Engineering, 11(1s), 275–284.

40) Sikder, M. Z., Shakil, M. A. I., Ahad, A., Karim, M. F., Intakhab, B., & Islam, D. A. (2025, June). Microwave-Based Detection of Early-Stage Renal Cell Carcinoma Using UHF Range Antenna. In 2025 International Conference on Computer Systems and Technologies (CompSysTech) (pp. 1-6). IEEE.

41) Ranjbareslamloo, S., Dzukey, G. A., Islam Muhit, M. M., & Qattawi, A. (2025). Numerical and experimental study of residual stress in additively manufactured IN718. Manufacturing Letters, 44, 915–927. https://doi.org/10.1016/j.mfglet.2025.06.108

42) Manoharan, D. (2025). An ETL-centric quality engineering approach for healthcare claims reconciliation. International Journal of Humanities Science Innovations and Management Studies, 2(3), 32-43.