International Research Journal of Multidisciplinary Scope (IRJMS), 2025; 6(1):171-177 IRJMS

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2025.v06i01.02407

Hybrid Contrastive Learning for Cyber Security Named Entry Recognition using Belief Rule Base

RadhaKrishna Karne[1]*, Kallem Niranjan Reddy[2], Kasapaka Rubenraju[3], Vijayalakshmi Chintamaneni[4], K Jamal[5]

[1,2]Department of ECE, CMR Institute of Technology, Hyderabad, Telangana, India, [3]Department of IT, Malla Reddy University, Hyderabad, Telangana, India, [4]Department of ECE, Vignan Institute of Technology and Science, Hyderabad, [5]Department of ECE, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, Telangana, India. *Corresponding Author's Email: krk.wgl@gmail.com

Abstract

It is possible to retrieve data from security-related occurrences, Named Entity Recognition (NER) is essential to cybersecurity. For rich contextual text embeddings, current approaches rely on pre-trained models; nevertheless, anisotropy presents a difficulty that may impact the quality of subsequent encoding. Furthermore, current models might have trouble detecting clutter. In order to deal with these problems, we provide a unique model that combines Belief Rule Base with Contrastive Learning for Named Entry Recognition (NER) in cybersecurity, called Hybrid Contrastive Learning (HCL), which is based on deep learning. Additionally, as a BRB parameter optimization method, the Distributed Constraint Covariance Matrix Adaptation Evolution Strategy (D-CMA-ES) is proposed. This paper contributes to advancing the state of the art in NER and provides insights into building more effective, interpretable, and scalable models for cybersecurity applications. Modelling and recognising entities across a wide range of cybersecurity data is crucial for effective and efficient response to cybersecurity crises. Neural networks are being used for entity extraction in the field of cybersecurity since Named Entity Recognition (NER) was developed. BRB used to improve the detection of fixed format entities is feasible and beneficial. As an alternative to the CMA-ES technique, we suggested the D-CMA-ES algorithm, which adaptively divides data into multiple subspaces for sampling, thereby mitigating the negative impact of high-dimensional samples on training outcomes. Experiments show that NER accuracy for cybersecurity is significantly improved when HCL is combined with the D-CMA-ES algorithm.

Keywords: Belief Rules Base, Cybersecurity, D-CMA-ES, NER.

Introduction

The growing frequency of cybercrimes and cyber-espionage occurrences has made cybersecurity more important for individuals, companies, and governments. When a cybersecurity incident occurs, analysts must quickly identify the involved parties from a variety of incident logs, which are compiled from data on host logs, cyber traffic, security alarms, and threat information. Although these entities are not readily visible in the real cyber circumstances, they have an impact on the cybersecurity scenario. Modelling and recognising entities across a wide range of cybersecurity data is crucial for effective and efficient response to cybersecurity crises. A thorough examination of the contextual influence on every word is made possible by these methods, which use encoders in the encoding phase or have been trained models throughout the representation process. Still, there are issues with cybersecurity data in NER. First of all, in vector space, embeddings produced by pre-trained language models, like BERT, frequently show excessive clustering and unequal distribution. This tendency may cause token sequencing that are semantically similar to be positioned farther apart, whereas tokens or sequences that are semantically unconnected might meet up with closely matched vectors. The model's capacity to correctly identify entities may be distorted by the inadequate representation of semantic similarity, which may also have an adverse effect on the model's overall performance by favouring particular directional biases. In addition, there is a deficit in the current approaches' ability to guarantee entity identification accuracy. Data related to cybersecurity appears to contain an enormous amount of noise. Despite this, existing models still classify these situations as IP addresses, indicating that their applicability has not been assessed. In this study, we present HCL, a rule the base of NER in cybersecurity that blends contrastive learning

(Received 13th September 2024; Accepted 14th January 2025; Published 27th January 2025)

with belief. Using the CMA-ES algorithm, our proposed approach, the Distributed CMA-ES (D-CMA-ES), searches among the several subspaces with comparatively smaller dimensions that are produced when the multivariate search space is partitioned. Because deep learning is developing so quickly, deep neural networks are now being considered as viable substitutes for conventional NER techniques (1). Rule-based and statistic machine techniques are the two main categories of early NER approaches. Expert-crafted rules that incorporate gazetteers and structural lexical characteristics are the foundation of rule-based approaches (2). Machine learning methods including SVM, CRF, HMM, and also the Perceptions are used in statistical procedures. The method for gathering information from internet vulnerability databases was created using machine learning as well as part-of-speech tagging (3). Despite very modest improvements in feature representation, a neural network architecture and training approach that lessens dependency on prior NLP expertise (4). By preserving context information from many time sequences, the model improves performance by using the linear stack of Bi-GRU with CNN to obtain hidden representations (5). More accurately, cybersecurity vulnerabilities are depicted by this updated word representation (6). In addition, an active learning approach based on advertisements that applies BiLSTM to word embedding encoding for network intrusion detection applications (7). An additional LSTM layer extracted adaptive cognitive hidden representations to construct pseudo-labels in order to solve the problem of data that were not correctly annotated (8). Without the need for preset information extraction sets, network threat data may be extracted from disordered APT reports using an open CyKG model and the neural Open Information Extraction (OIE) technique based on attention mechanisms (9). The open-source Python application CyNER extracts cyber-security-related components and indications of compromise (IOC) using heuristic techniques and transformer-based models (10). The technology provides robust adaptability and several trained models (11).

Methodology

First, we give a brief overview of HCL and present a framework example in Figure 1. First, BERT is used to convert sentences into embedding matrices. To improve BERT in this process, we use contrastive learning (12). In particular, we acquire span representations for every sentence item. We then use a mini-batch to generate models of the span components, the final symbol, and the first symbol for the same kind of entity. Using contrastive learning, we suggest three goals for NER based on it (13). Lastly, we employ BiLSM to increase bi-directional semantic dependency capture across long distances by splicing forward and backward concealing vectors (14). The Multihead Self attention layer, sometimes known as the MS layer, is also added, allowing HCL to selectively focus on higher-value segments of the input sequence even when noise is present. Subsequently the possibility that each token will belong to a different label is predicted using the CRF model (15). In the end, a BRB is used to filter away entities that are incorrectly recognized; improving the accuracy of cybersecurity entity recognition, especially for entities typed using fixed-format phrases that lack semantics. Figure 1 represents the Hybrid Contrastive Learning model (16).

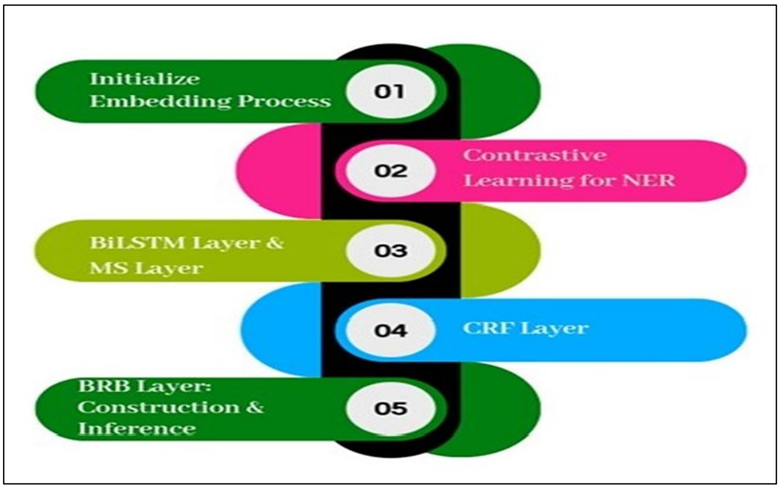
Initialize Embedding Process
01
02
Contrastive Learning for NER
BiLSTM Layer & MS Layer
03
04
CRF Layer
BRB Layer: Construction & Inference
05

Figure 1: Proposed Hybrid Contrastive Learning Model

Initialize Embedding

HCL first converts each token in the text into a vector v with embedded words and position embedding using BERT.

$$V = (V_W, V_P) \qquad [1]$$

Vector is a concatenation of word embedding and position embedding.

NER Learning Model

To fine-tune the BERT, we first apply contrastive learning after acquiring the initial token vector. Contrastive learning is mostly utilized in representation learning to lessen the numerous BERT quirks. Its main objective is to move the vector space embedded data of dissimilar texts apart and those of comparable works closer together. We may create a vector image for an on-going series of symbols in a particular sentence by placing a beginning token in position x and a final symbol in place y.

$$Span_{x,y} = Lll[V_x \oplus V_y \oplus l(y-x)] \qquad [2]$$

Where Lll is learnable linear layer

$\oplus$ is vector concatenation

l(y-x) is (y-x)th row of a learnable span width embedding matrix

where span $_{x,y}$ indicates a sequence of vector illustration of an entity.

Every non-entity token sequence is penalized equally by the span-based objective. Therefore, we present a position-based object to determine the limits of token sequences that reflect a certain thing. It makes logical to place the first and last token of the same type closer to the embedding space. More precisely, in a token layout of objects of the same type in a mini-batch, we identify the first version using the first and last symbol as follows.

$$B_z^{start} = \frac{\sum_{x=1}^{z} \quad span_{0,0}^{x}}{z} \qquad [3]$$

$$B_z^{stop} = \frac{\sum_{x=1}^{z} \quad span_{n,n}^{x}}{z} \qquad [4]$$

Where n is the number of token in span^x

BiLSTM layer

In this section, the sentence matrix is encoded using BiLSTM. First, we compute a forgetting gate to decide what data to throw away. Second, we determine which data should be memorized by computing the memory gate. Additionally, we compute the current cell state along with the temporary and previous cell states to integrate the process of forgetting and memory gates. Finally, we find the output gate and the hidden layer state.

MS layer

We employ the MS layer to capture the sequence-wise dependency across tokens and enhance the resilience of HCL after BiLSTM has finished encoding the embedding. The similarity of each value token to each key token and query token determines its weight. The input is allocated to h different subspaces using a parameter matrix. The scaled dot product concentration score is calculated for each one individually, and the result is combined as the final attention score. This method uses a hidden layer length (h) time to obtain the MS score.

$$u_i = softmax(\frac{Q_1 K_I^T}{\sqrt{D}})T \qquad [5]$$

$$MS = (u_1, u_2, \ldots\ldots, u_h) \qquad [6]$$

CRF layer

We see the cybersecurity entity extraction process in this layer as a series of labeling operations. The n × m vector P, where n is the total number of data tokens and m is the number of label types, is produced after the MS layer processes the data. The input indicates the probability that the token y's label x will appear higher in the phrase.

BRB layer

The BRB layer has an internal structure, which sets it apart from the previously stated data-driven models. Furthermore, the BRB model can explain all types of ambiguous information and fully exploit semi-quantitative information, in contrast to the previously described data-driven models. We design a BRB that consists of several rules. One of the foundational characteristics in these rules is the CRF output; as an extra criterion, we apply robust and easily comprehensible regexes. Every rule has a primary attribute with a defined weight, and to indicate how credible the conclusion is, confidence is matched with the rule's last section. The BRB model is explained in the following manner.

$$K_r\text{: for } (i_1 \text{ is } A_1^k) \wedge (i_2 \text{ is } A_2^k) \ldots\ldots. (i_n \text{ is } A_n^k)\text{,then } (D_1,\beta_1)\ldots\ldots.(D_2,\beta_2) \qquad [7]$$

Where ∧ designates intersection

n_i represents the ith attribute of BRB model, and n number of attributes

K_r is rth rule of BRB model,

A_n^k is reference value of nth premise attribute in kth rule.

Figure 2 illustrates the BRB model's structure. The input will make the matching rules active in accordance with the BRB model. The Evidential Reasoning (ER) technique will then be used to integrate the activation rules and produce the inference results. The D-CMA-ES algorithm was used to solve the high-dimensional, restricted optimization issue. This method uses the CMA-ES algorithm to search in the low-dimensional subspaces after dividing the multidimensional search space into many subspaces with much reduced dimensions. Then, by combining the findings from every search, the original problem's solution is discovered.

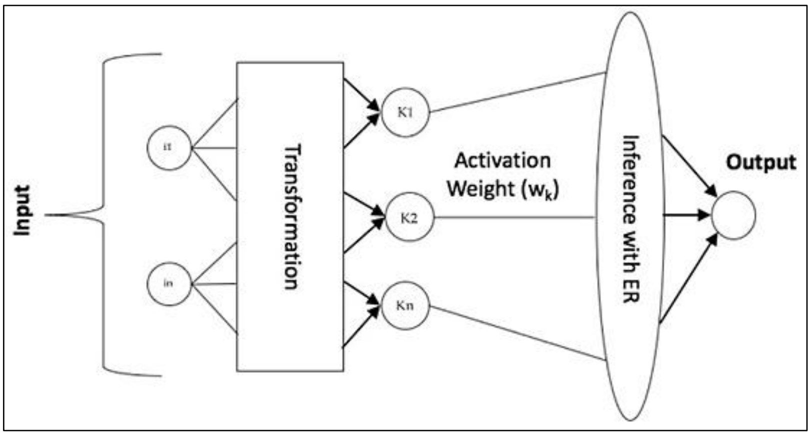

Figure 2: Basic structure of BRB System

Results and Discussion

The performance of several models on our gathered dataset Open CS is shown in this section. We gather and compile a substantial amount of free and unstructured data Open CS. The most recent APT reports, Amazon's network security blog, Treat Intelligence of Client Vault, and CVE vulnerability description entries are a few of the data sources. Based on this, ten entity categories—organization (ORG), software (SOF), malware (MAL), vulnerability (VUL), identity (IDE), tool (TOO), protocol (PRO), system (SYS), equipment (EQU), and attack techniques (MET)—are ultimately developed and used to name entities in the cybersecurity data. The UCO ontology's entity categories, word frequency data, and the analytical outcomes of the cybersecurity data that has been filtered are the foundations for these categories. Here is how we contrast HCL with four baseline models. Initially, the CRF model is a statistically based conditional probability distribution that is often applied to sequence tagging. Second model is the BiLSTM-CRF architecture for the NER challenge on cybersecurity. After the BiLSTM layer uses input embedding to extract contextual characteristics, the CRF layer decodes sequences in order to predict labels. Third one is a linear combination of CNN and LSTM at the deep neural network layer to represent global and local features more effectively. Lastly, a module for entity border detection for the prediction of entity head and tails (GNNs & RNNs).

Table1: Performance comparison of different models on open CS (MalNet) dataset

Model	P%	R%	F1%
CRF	79.35	71.59	75.27
BiLSTM-CRF	85.64	84.19	84.9
Linear Stack LSTM	80.71	78.92	79.80
GNN & RNN	91.06	90.13	90.59
HCL with BRB	92.63	89.36	89.96

Precision (P), Recall (R), and F1—all commonly used evaluation metrics in information extraction tasks—are utilized to analyze the model's performance. P and R, which stand for the percentages of correct and incorrect samples, respectively, that the model correctly detected in all identified samples, are shown in Table 1. The model's overall performance is assessed using the F1 value, which is the harmonic average of accuracy and recall. Figure 3 is the graphical representation of comparison of proposed with various models with respect to the evaluation parameters.

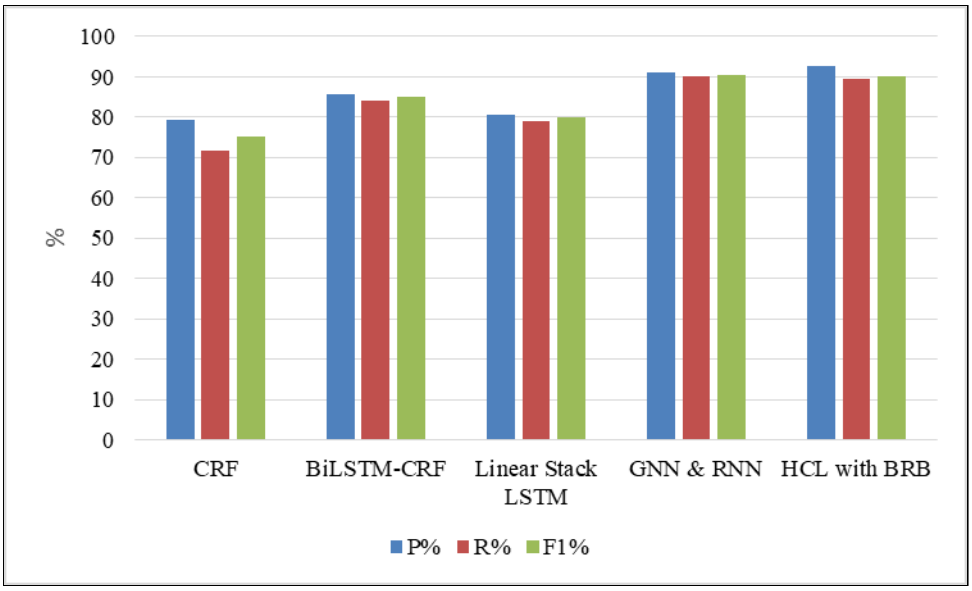

Figure 3: Performance of Various Entities in Various Models

Figure 4 displays an average correlation between the token sequence representations and the sample for batch positives and negatives for span-based contrastive learning. We find that the degree of similarity of in-batch negatives rapidly decreases with increasing training, indicating that in-batch negatives provide only a small number of gradient signals. However, our strategy successfully increases the similarities of token sequence representation for the same sort of vector space entities, as the resemblance for in-batch positives remains high and can be easily separated from the negatives. In position-based contrastive learning, Figure 5 shows the average similarity between token sequence representations and sample in-batch positives and negatives. Using BRB to improve the detection of fixed format entities is both feasible and beneficial. We proposed the D-CMA-ES algorithm as a substitute for the CMA-ES approach, which adaptively splits data into several subspaces for sampling, reducing the detrimental effect of high-dimensional samples on training results.

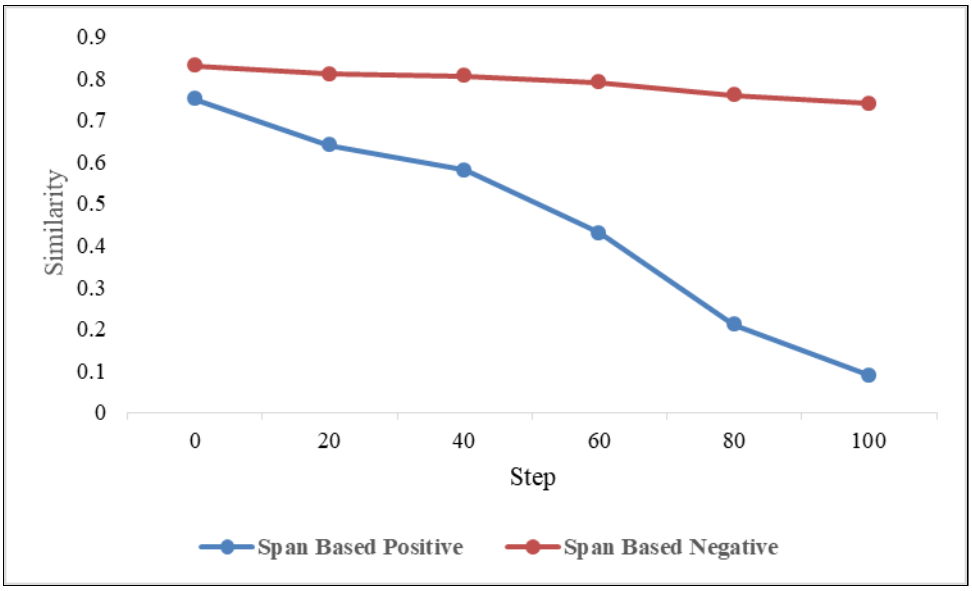

Figure 4: Similarity Representation of Span Based Models

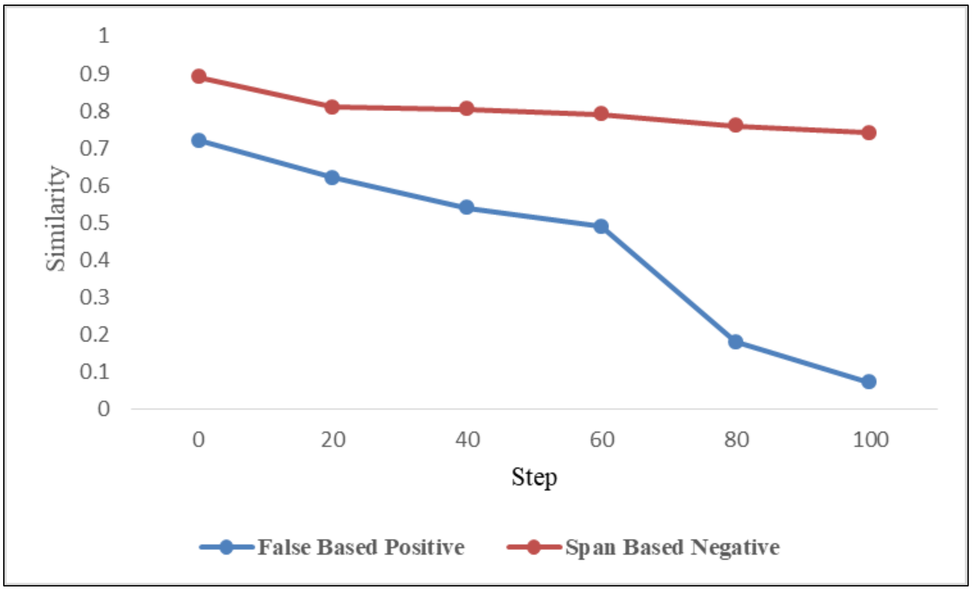

Figure 5: Similarity Representation of Position and Span Based Models

Conclusion

In this work, we introduce HCL with BRB, a novel NER model for cybersecurity. By using contrastive learning, HCL modifies BERT to create goals based on location and span. This technique lessens the effect of anisotropy on encoding quality by increasing the vector space comparability of token sequence representations for entities of the same class. We also show that using BRB to improve the detection of static format entities is feasible and beneficial. We suggest the D-CMA-ES algorithm as an alternative to the CMA-ES approach, which effectively reduces the detrimental effect of high-dimensional samples on training results by adaptively dividing data into several subspaces for sampling. Experimental evaluations demonstrate the benefits of JCLB for NER in cybersecurity.

Abbreviation

Nil.

Acknowledgement

Authors are thankful to CMR Institute of Technology for providing literature collection facilities. All authors listed have significantly contributed to the development and the writing of this article.

Author Contributions

All authors have been personally and actively involved in substantial work leading to the paper and will take public responsibility for its content. RadhaKrishna Karne: Conceptualization, Methodology, Data Curation, Writing- Original Draft, Writing Review and Editing. Kallem Niranjan Reddy: Formal Analysis, Writing – Review and Editing. K Ruben Raju: Validation, Supervision, Resources, Writing -Review and Editing. Vijayalakshmi: Visualization, Project Administration, K Jamal: Methodology, Data Analysis, Writing – Review & Editing.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics Approval

Not applicable.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

References

1. Liu X, Tian J, Niu N, Li J, Han J. “Standard Text” Relational Classification Model Based on Concatenated Word Vector Attention and Feature Concatenation. Applied Sciences. 2023 Jun 14;13(12):7119.
2. Zacharis A, Patsakis AiCEF: an AI-assisted cyber exercise content generation framework using named entity recognition. C. International Journal of Information Security. 2023 Oct;22(5):1333-54.
3. Wang X, Liu J. A novel feature integration and entity boundary detection for named entity recognition in cybersecurity. Knowledge-Based Systems. 2023 Jan 25;260:110114.
4. Lughbi H, Mars M, Almotairi K. CybAttT: A Dataset of Cyberattack News Tweets for Enhanced Threat Intelligence. Data. 2024 Feb 23;9(3):39.

5. Silalahi S, Ahmad T, Studiawan H. Transformer-based named entity recognition on drone flight logs to support forensic investigation. IEEE Access. 2023 Jan 5;11:3257-74.
6. Hu C, Wu T, Liu C, Chang C. Joint contrastive learning and belief rule base for named entity recognition in cybersecurity. Cybersecurity. 2024 Apr 3;7(1):19.
7. Karne RK, Sreeja TK. PMLC-Predictions of Mobility and Transmission in a Lane-Based Cluster VANET Validated on Machine Learning. International Journal on Recent and Innovation Trends in Computing and Communication. 2023;11:477-83.
8. Chen J, Lu Y, Zhang Y, Huang F, Qin J. A management knowledge graph approach for critical infrastructure protection: Ontology design, information extraction and relation prediction. International Journal of Critical Infrastructure Protection. 2023 Dec 1;43: 100634.
9. Miao P, Du Z, Zhang J. DebCSE: Rethinking Unsupervised Contrastive Sentence Embedding Learning in the Debiasing Perspective. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management 2023 Oct 21 (pp. 1847-1856).
10. Karne RK, Sreeja TK. A Novel Approach for Dynamic Stable Clustering in VANET Using Deep Learning (LSTM) Model. IJEER. 2022;10(4):1092-8.
11. Esho AO, Iluyomade TD, Olatunde TM, Igbinenikaro OP. A comprehensive review of energy-efficient design in satellite communication systems. International Journal of Engineering Research Updates. 2024;6(02):013-25.
12. Karne RK, Sreeja TK. Cluster based vanet communication for reliable data transmission. InAIP Conference Proceedings 2023 Nov 21 (Vol. 2587, No. 1). AIP Publishing.
13. Alhamyani R, Alshammari M. Machine Learning-Driven Detection of Cross-Site Scripting Attacks. Information. 2024 Jul 20;15(7):420.
14. Al-Baity HH. The artificial intelligence revolution in digital finance in Saudi Arabia: a comprehensive review and proposed framework. Sustainability. 2023 Sep 15;15(18):13725.
15. Mishra S. Exploring the impact of AI-based cyber security financial sector management. Applied Sciences. 2023 May 10;13(10):5875.
16. Jia Y, Gu Z, Du L, Long Y, Wang Y, Li J, Zhang Y. Artificial intelligence enabled cyber security defense for smart cities: A novel attack detection framework based on the MDATA model. Knowledge-Based Systems. 2023 Sep 27;276:110781.