IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 February 2025, accepted 3 March 2025, date of publication 11 March 2025, date of current version 24 March 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3550378

RESEARCH ARTICLE

Leveraging Cyberattack News Tweets for Advanced Threat Detection and Classification Using Ensemble of Deep Learning Models With Wolverine Optimization Algorithm

SRIPADA NSVSC RAMESH[1], BADER MOHAMMED M. AL FARDAN[2,3], C. S. S. ANUPAMA[4], KOLLATI VIJAYA KUMAR[5], SEONGSOO CHO[6], SRIJANA ACHARYA[6], AND CHEOLHEE YOON[7]

[1]Department of CSE, Aditya College of Engineering and Technology, Surampalem 533437, India
[2]Department of Industrial Engineering, College of Engineering, King Khalid University, Abha 61421, Saudi Arabia
[3]Center for Engineering and Technology Innovations, King Khalid University, Abha 61421, Saudi Arabia
[4]Department of Electronics and Instrumentation Engineering, V. R. Siddhartha Engineering College, Deemed to be University, Vijayawada 520007, India
[5]Department of Computer Science Engineering, GITAM School of Technology, GITAM University, Visakhapatnam 530045, India
[6]Department of Convergence Science, Kongju National University, Gongju 32588, South Korea
[7]Police Autonomous Driving Laboratory, Korean National Police University, Asan 31539, Republic of Korea

Corresponding authors: Srijana Acharya (srijana@kongju.ac.kr) and Cheolhee Yoon (bertter@police.ac.kr)

This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant funded by Korea Government (MSIT) (Development of causal reasoning and expression technology to enhance the proof of digital evidence) under Grant RS-2023-00225661.

ABSTRACT At present, cyber-attacks have become more critical and familiar, which appeals to a novel line of security defences to defend against them. Cyber Threat Intelligence (CTI) originated as a reputation in the frequently developing cybersecurity landscape, vital in safeguarding digital models. Understanding this domain develops over a complete study of intelligence's numerous features and sources. Data sharing and analysis centres perform as alarms of collaboration, demonstrating the collective vigilance needed to oppose such developing attacks. Inspecting and gathering data about cyberattacks from tweets can effectively deliver critical perceptions of the threats, their effects, occurrence areas, and probable mitigation tactics. Existing study on cyberattack absences in establishing Artificial Intelligence (AI) based analytic solutions for delivering country-wide cyber-attack intelligence. Cyber planners at a domestic level need AI-based decision support models to determine a country's cyber attitude or vigilance. This study designs and develops an Enhanced Threat Intelligence for Cybersecurity Using an Ensemble of Deep Leaning Models with Metaheuristic Optimization Algorithm (ETIC-EDLMOA) model. The presented ETIC-EDLMOA model's main aim is to detect and mitigate network attacks in cybersecurity effectively. Initially, the ETIC-EDLMOA model undergoes a data pre-processing stage to ensure clean and structured input data for analysis. Besides, the Word2vec model is utilized for feature extraction. For the classification process, the ensemble of DL models is employed, including the recurrent neural network (RNN) method, long short-term memory (LSTM) model, and conditional variational autoencoders (CVAE) technique. Finally, the ensemble models' hyperparameter fine-tuning process is performed using the Wolverine optimization algorithm (WoOA) technique. A comprehensive range of simulation analyses is conducted to ensure the improved performance of the ETIC-EDLMOA method on the CybAttT dataset. The comparison study of the ETIC-EDLMOA method illustrated a superior accuracy value of 98.51% over existing techniques.

INDEX TERMS Cyber threat intelligence, deep leaning, wolverine optimization algorithm, cybersecurity, Word2vec.

The associate editor coordinating the review of this manuscript and approving it for publication was S. K. Hafizul Islam.

I. INTRODUCTION

Cyber-attacks are continuously rising in complexity and frequency. Cybercriminals can bypass organizational control

© 2025 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

securities using customized campaigns and intrusion kill chains, tactics, techniques, and procedures (TTPs) [1]. Cybersecurity outages and breaches have been extensively protected in the media, and statistical problems regarding the number of cyber-attacks are obtainable from several resources. Even though several cybersecurity breaches, there is little knowledge of the investigation of the regions, which organizations would prioritize to increase their efficiency in addressing identified attacks while reducing the hazard from developing threats [2]. One of these techniques can assist in lowering security breaches by applying and emerging strong CTI. CTI examines tendencies and technological advances in 3 regions: Hacktivism, Cyber Espionage, and CS. Countries utilize CTI as an effective solution to develop preventive cybersecurity events in development and as an outcome to support international security [3]. CTI is a part of cybersecurity that involves the contextual data encompassing cyber threats, which is the knowledge of the present, future, and past TTPs of a broad range of threat actors. It is timely and actionable and holds business values, and it can update the security squads in adversarial organization entities so that they can evade them [4].

CTI is also a proactive security measure that involves collating, analyzing, and gathering information regarding possible threats in the real world to prevent adverse results and data breaches. Usually, advanced persistent threats (APT) are highly targeted and sophisticated forms of cyber threat. APT signifies the most advanced kinds of threats opposing innovative systems, which are verified to be highly complex to combat [5]. Attackers employ advanced threat methods to remotely control affected machines and exfiltrate complex information from governments and organizations. Owing to the dynamic nature of the APT threat process, undertaking systems using security products that depend on classical defences often fails to identify APT corruptions. Moreover, states support cyber warfare and the rise of structured cybercrime organizations to verify additional complexities in the digital world [6]. Fig. 1 portrays the structure of enhanced threat intelligence for cyberattacks. Artificial Intelligence (AI) is altering the area of cybersecurity by allowing more automated and advanced attack investigation, recognition, and response. AI methods, driven by Machine Learning (ML) models, can identify complex patterns, detect anomalies, and process vast numbers of data, which could specify possible cyber-attacks. Different classical security approaches that depend on signature-based recognition and pre-defined rules, AI can change to novel attacks and eventually constantly enhance its execution [7].

In the cyber-attack intelligence framework, AI is vital in moving from defence approaches to proactive attack forecasts. By examining network traffic, user behaviour, historical data, and other inputs, AI-driven methods can predict possible threats, detect developing attack actors, and offer real-world references to reduce risks [8]. AI also improves vulnerability management, incident response, and attack hunting, making cybersecurity more effective and efficient in quickly developing attack landscapes. The growing advancement and frequency of cyber-attacks present a crucial challenge to organizations striving to protect their systems and data. Conventional cybersecurity methods often fall short in identifying and preventing evolving threats due to the rapidly changing tactics employed by cybercriminals [9]. A key to improving defence mechanisms is utilizing advanced analytical techniques to detect and classify potential threats early on. Organizations can gain valuable insights into attack trends and emerging risks by analyzing real-time data, such as cyberattack-related news. Using cutting-edge ML models and optimization algorithms can significantly improve the capability to predict and prevent future cyber threats, making proactive threat detection a critical step in improving overall cybersecurity [10].

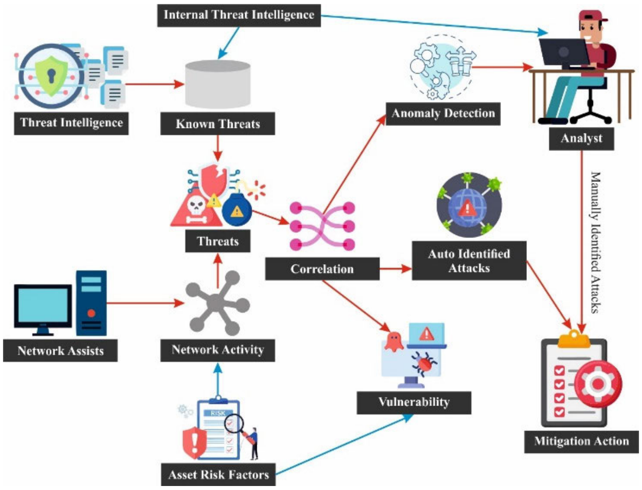

FIGURE 1. **The architecture of enhanced threat intelligence for cybersecurity.**

This study designs and develops an Enhanced Threat Intelligence for Cybersecurity Using an Ensemble of Deep Leaning Models with Metaheuristic Optimization Algorithm (ETIC-EDLMOA) model. The presented ETIC-EDLMOA model's main aim is to detect and mitigate network attacks in cybersecurity effectively. Initially, the ETIC-EDLMOA model undergoes a data pre-processing stage to ensure clean and structured input data for analysis. Besides, the Word2vec model is utilized for feature extraction. For the classification process, the ensemble of DL models is employed, including the recurrent neural network (RNN) method, long short-term memory (LSTM) model, and conditional variational autoencoders (CVAE) technique. Finally, the ensemble models' hyperparameter fine-tuning process is performed using the Wolverine optimization algorithm (WoOA) technique. A comprehensive range of simulation analyses is conducted to ensure the improved performance of the ETIC-EDLMOA method on the CybAttT dataset. The key contribution of the ETIC-EDLMOA method is listed below.

- The ETIC-EDLMOA model begins with a thorough pre-processing phase that optimizes data for analysis.

This step ensures that relevant features are extracted and noise is minimized, improving the overall efficiency and effectiveness of the model in subsequent stages.

- The ETIC-EDLMOA technique employs Word2Vec-based feature extraction to transform textual data into dense, meaningful vector representations. This technique captures semantic relationships between words, enabling the model to comprehend the underlying patterns better. This significantly improves the method's capability to process and classify textual information.
- The ETIC-EDLMOA methodology utilizes an ensemble of DL techniques, namely RNN, LSTM, and CVAE, to improve classification accuracy. Each model presents unique strengths, capturing diverse aspects of the data. This incorporation enables the model to handle complex patterns effectively in the classification task.
- The ETIC-EDLMOA approach implements the WoOA model to fine-tune the method's parameters, ensuring optimal performance across classification tasks. Searching for the most effective parameter set improves the method's accuracy and efficiency. This optimization process results in enhanced generalization and improved results on the dataset.
- Integrating an ensemble of advanced DL models with the WoOA for fine-tuning presents a novel approach. This integration allows the model to employ the merits of each technique while optimizing parameters for maximum performance. The novelty is using WoOA to fine-tune DL models, improving the classification process's robustness and accuracy on the CybAttT dataset. This integrated approach significantly outperforms traditional methods.

II. LITERATURE SURVEY

In [11], CTIMD presents a DL-based dynamic malware recognition approach that incorporates understanding Cyber Threat Intelligence (CTI) threats into learning API call sequences with run-time parameters. It removes Indicators of Compromise (IOCs) from CTIs and utilizes IOCs to help the security-sensitive recognition stages of API calls. Lastly, it sustains the characteristics vector classifications into Deep neural networks (DNNs) to train the malware recognition method. Yu et al. [12] present CNN methods that depend on hierarchical knowledge attention and migration mechanism, called CNN depends on hierarchical knowledge attention and migration mechanism (HM-ACNN). The primary stage in HM-ACNN is changing the CTI text into a dual-dimension image, depending on the embedded method. The tactic-to-technique understanding of movements to be accomplished by altering the parameters of the attention and CNN layers in the tactic's categorization process depends on the distinctive hierarchical relationship between techniques and tactics. Afterwards, the method classification is ended by fine-tuning. Ahmed et al. [13] aim to efficiently remove cyber-related triplicates by utilizing the joint removal methods that solve the problems in the traditional pipeline method. Initially, the BIEOS tagging system was employed to CTI data utilizing the joint tagging method, and then the relation trios were mutually removed. At last, the related trios were removed utilizing the related-matching method, which was subsequently identical to the ideally suited relation for the dual predicted objects. Sufi [14] presents an AI-based solution that individually gathers multi-dimensional cyber-threat information on social media posts on cyber-related outcry. The projected method offers significant analytical abilities in the cyber-threat range and utilizes advanced AI-based models for forecasting, recognition of abnormality, location recognition, translation, sentiment analysis, and more. Additionally, the method achieved AI-based acquisition.

Azar et al. [15] present Enhanced Metaheuristics with Hierarchical DL-based Attack Detection (EMHDL-AD) approach in a cloud-based CPS framework. The presented EMHDL-AD approach recognizes several kinds of threats to protect CPS. In the primary phase, data pre-processing is applied to change the input dataset into a suitable format. Afterwards, the Quantum Harris Hawks Optimization (QHHO) model is utilized for feature selection. An Improved Salp Swarm Algorithm (ISSA) is utilized to improve the hyper-parameters of the HDL methods to identify various threats. In [16], a method with a structure is presented to determine malware threats by utilizing AI approaches to protect distributed and different situations. This innovative approach enables the proactive tracking of system traffic information to identify threats and malware in the IoT ecosystem. Furthermore, the new method creates smarter environments that are more stable and aware of probable future attacks. Zhang et al. [17] associate CTI with organization Security Requirements (SR) data to build a Knowledge Graph (KG) called RCTI and forecast novel understanding on the diverse graph. EGNN, a new graph neural networks (GNNs)-based method, is projected to describe edges by demonstrating an advanced model for generating the data edges. At last, the EGNN method is utilized to forecast novel links on the RCTI graph. Chen et al. [18] present an innovative cyber-attack intelligence removal method named Cyber Attack Relation Extraction (CARE). It removes significant attack entities and offers their relationship in both textual and graphical types, which assist cybersecurity staff in quickly grasping the main information from security reports. To take attack-related data, this research accepts BERT to improve contextualized word representation and employs Transfer Learning (TL) to remove the relations among attack entities. Al Mamun et al. [19] introduce Feature Evolution using Genetic Programming (FEGP) for Advanced Persistent Threat (APT) detection. FEGP improves APT detection by automatically constructing discriminative features using multi-tree Genetic Programming (GP). The method also employs SHapley Additive exPlanations (SHAP) and Gaussian Naive Bayes (GNB) for its processes.

Sharaf et al. [20] propose the Mountain Gazelle Optimization with Deep Ensemble Learning based intrusion detection

(MGODEL-ID) technique using MGO for feature selection and an ensemble of LSTM, deep autoencoder (DAE), and extreme learning machine (ELM) classifiers. Hyperparameters are tuned with the Dung Beetle Optimizer (DBO). Kalutharage et al. [21] present an AI-based anomaly detection model using an explainable AI (XAI) technique. The model enhances detection accuracy by incorporating domain knowledge and the MITRE ATT&CK framework. Baluguri et al. [22] propose a stacked ensemble method integrating Decision Trees (DTs), k-nearest neighbour (KNN), multi-layer perceptron (MLP), and NB with a Logistic Regression (LR) meta-learner for improved network intrusion detection. Alrayes et al. [23] introduce the Privacy-Preserving Statistical Learning with an Optimization Algorithm for a High-Dimensional Big Data Environment (PPSLOA-HDBDE) approach ensures privacy-preserving statistical learning in high-dimensional big data, using feature selection, an ensemble of temporal convolutional network (TCN), multi-layer auto-encoder (MAE), and Extreme Gradient Boosting (XGBoost) models for intrusion detection, and hyperparameter tuning with the improved marine predator algorithm (IMPA) method. Batchu et al. [24] present an intrusion detection system using DL, with data pre-processing, balancing via CGAN, and classification using a stacked sparse denoising autoencoder (SSDAE) optimized by the firefly-black widow (FA-BW) hybrid algorithm. Antonijevic et al. [25] present a hybrid framework combining CNN, Categorical Boosting (CatBoost), and Light Gradient-Boosting Machine (LightGBM) models, optimized by metaheuristic algorithms, for efficient attack detection and classification in IoT networks. Kumari et al. [26] optimize Artificial Neural Network (ANN) layers using Spider Monkey Optimization (SMO) to detect network intrusions and attacks in real time. Papalkar and Alvi [27] propose a Creative Swagger (CS) Optimized Deep-CNN for detecting and mitigating DDoS attacks, using the CS algorithm to tune Deep CNN parameters for improved accuracy, with initial verification through a exclusion list table. Oyinloye et al. [28] present an updated learning strategy for ANN to address data categorization issues caused by unbalanced data in intrusion detection systems. Table 1 summarizes the existing studies on cyberattack detection using deep learning ensemble.

Despite the enhancements in IDS, several challenges remain. Many existing methods face difficulty with data imbalance, resulting in biased detection outputs, and the reliance on specific algorithms often limits their adaptability to new attack patterns. Furthermore, while optimization techniques enhance model performance, they may still fail to handle large-scale, dynamic environments like IoT or cloud systems. The integration of domain knowledge in some models is still in its infancy, and their interpretability can be limited. Many approaches also concentrate on improving detection accuracy without adequately addressing real-time scalability and robustness against novel, advanced cyber-attacks. Additionally, the requirement for more comprehensive evaluation across diverse datasets and threat scenarios persists.

III. MATERIALS AND METHODS

This study proposes the ETIC-EDLMOA model. The presented model aims to detect and mitigate network attacks in cybersecurity effectively. The model comprises stages such as data pre-processing, feature extraction, ensemble classification process, and parameter optimizer to achieve that. Fig. 2 demonstrates the entire flow of the ETIC-EDLMOA method.

A. STAGE I: DATA PRE-PROCESSING

Initially, the ETIC-EDLMOA model undergoes a data pre-processing stage to ensure clean and structured input data for analysis. Data pre-processing involves converting raw data into an organized structure appropriate for analysis [29]. The below-mentioned are pre-processing steps employed to the gathered data:

- Cleaning every emoji and pronouns utilizing a Python package of emoji.
- Eliminating stop words, which contain no values inserted into the developed method. These words were described in the Natural Language Toolkit (NLTK).
- Exclusion of punctuations: Punctuation like (~¦+|!--,;) are eliminated
- Repeated character elimination: characters like (“…”, “//”), hyphens, brackets, and symbols are eliminated.
- Links and mentions removal: all links and mentions that begin with @|https were deleted.

There is a severe divergence in the spread of tweet lengths (in characters) after and before the pre-processing phase for the complete database. There is a substantial decrease in the distance of the mainstream of tweets, which changes from 500 characters or fewer earlier pre-processing to 300 characters or fewer after pre-processing.

B. STAGE II: WORD2VEC MODEL

Besides, the Word2vec model is utilized to extract features [30]. This model is chosen because it can learn distributed word representations in a continuous vector space. The key merit of Word2Vec is its ability to capture semantic meaning and contextual relationships between words based on their usage in a corpus, making it superior to traditional models like Bag-of-Words (BoW) or TF-IDF, which only consider word frequency and disregard context. Unlike these older techniques, Word2Vec accounts for word proximity and co-occurrence, resulting in dense vectors better capturing similarities and relationships between words. Furthermore, Word2Vec is computationally effective, allowing for large-scale learning from massive datasets. Its capacity to generate high-quality word embeddings also enables its use in downstream tasks such as classification, clustering, and sentiment analysis. These factors make Word2Vec specifically advantageous when modelling complex textual data in real-world applications. Fig. 3 depicts the Word2Vec framework.

TABLE 1. Summary of existing studies on cyberattack detection using deep learning ensemble.

Ref	Techniques	Metrics	Findings
[11]	DL, CTI, IOC, DNN	F1-score, Accuracy	CTIMD outperforms existing methods, improving F1-score by 4.0%–41.3% on raw API calls and 1.2%–6.5% on API calls with run-time parameters.
[12]	CNN, Hierarchical knowledge migration, Attention mechanism, Word embedding model, Tactic-to-technique knowledge migration	F1-score for tactics, F1-score for techniques	HM-ACNN outperforms CNN, RNN, and CRNN, with F1 Scores of 93.66% and 86.29% for tactics and techniques classification, respectively.
[13]	BIEOS tagging scheme, Joint tagging technique, Attention-based RoBERTa-BiGRU-CRF model, Relation-matching technique	Precision, Recall, F1-score	The proposed model achieved a 7% improvement in F1-score, outperforming existing models in CTI knowledge triple extraction.
[14]	AI-based algorithms, CNN, Sentiment analysis, Location detection	30,203 records collected, 3,789 cyber-related tweets analyzed, 893 non-English tweets translated	The system autonomously detected anomalies, predicted cyber-attacks, and provided evidence-based decisions across multiple platforms.
[15]	QHHO, ISSA, HDL	Benchmark intrusion datasets	The EMHDL-AD method outperformed existing approaches in detecting cyber-physical attacks in CPS environments.
[16]	DNN, AI, ML	Detection accuracy, F1-score, Network bandwidth increase, CPU consumption increase, Memory usage, Power consumption increase	The proposed methodology efficiently detects malware and attacks in IoT smart environments with high accuracy and minimal resource impact.
[17]	CTI, KG, GNN, EGNN	Connectivity rate, 3 public datasets	The EGNN model effectually predicts new links in the RCTI graph, improving cybersecurity detection and management in critical infrastructure.
[18]	CTI, BERT, TL, Relation Extraction	F1-score	The CARE technique achieves a 97% F1 Score, effectively extracting and presenting cyber threat relations for improved decision-making in cybersecurity.
[19]	GP, Feature Evolution, Multi-tree GP, APT Detection	Bal_acc, F1_score, TNR, FPR	FEGP enhances APT detection accuracy, outperforming comparative methods by up to 3.73%, illustrating the efficacy of GP-based feature construction.
[20]	MGO, LSTM, DAE, ELM, IDS, DBO, Z-score Normalization	Accuracy, Precision, Recall, F-score, G-measure	MGODEL-ID significantly outperforms other models in intrusion detection, depicting improved security for smart grid environments.

TABLE 1. *(Continued.)* **Summary of existing studies on cyberattack detection using deep learning ensemble.**

[21]	Anomaly Detection Model, AI, KG Integration, XAI, MITRE ATT&CK Framework	Detection Accuracy, Threat Intelligence Accuracy, Attack Legitimacy Validation	The presented model achieved high detection accuracy (0.97) and provided 100% accuracy for threat intelligence, improving IoT security with actionable insights and faster responses.
[22]	Stacked Ensemble Method, DT, KNN, MLP, NB, LR Meta-Learner, Feature Engineering, Data Pre-processing	Accuracy, Precision, Recall, F1-score	The stacked ensemble model achieved high output, efficiently detecting network intrusions and offering a scalable, flexible solution for cyberattack prevention.
[23]	ML, LSN, SCSO, FS, TCN, MAE, XGBoost, IMPA, Ensemble Learning	Accuracy, Precision, Recall, F1-score, MCC	The PPSLOA-HDBDE approach achieved a superior accuracy of 99.49% in intrusion detection while ensuring privacy preservation in high-dimensional big data environments.
[24]	DL, Data Pre-processing, Data Balancing, CGAN, SSDAE, FA-BW, Hybrid Optimization, IDS	Accuracy, Precision, Recall, F1-score, AUC	The proposed framework using DL and hybrid optimization significantly improved DDoS attack detection accuracy compared to existing methods.
[25]	CNN, ML, CatBoost, LightGBM, Metaheuristics Optimizers, Hybrid Framework, XAI	Accuracy, Precision, Sensitivity, and F1-score	The proposed hybrid framework attained 99.83% accuracy in classifying IoT network attacks and improved security with explainable AI insights.
[26]	ANN, SMO, IDS	Accuracy, Precision, Recall, F1-score	For attack detection, the SMO-ANN model achieved 100% accuracy on the Luflow dataset and 99% accuracy on the NSL-KDD dataset.
[27]	CS Optimization, DeepCNN	Accuracy, Sensitivity, Specificity	The CS-optimized DeepCNN approach achieved high efficiency in detecting DDoS attacks with 97.07% accuracy, 97.23% sensitivity, and 96.91% specificity on the UNSW-NB15 dataset.
[28]	ANN, Random Weight Augmentation, Standard Scaler	Accuracy, Precision, Sensitivity, and F1-measure	The augmented ANN achieved a significant 92% accuracy, enhancing resilience to disturbances and computational complexity in intrusion detection systems.

Word2vec is one of the NN probabilistic methods for making word vectors. It is a primary shallow neural network that utilizes a sliding window to state secure-size contextual and tries to forecast the contextual for the present term or to utilize the contextual to forecast the present word. This model includes dual training methods, such as the Skip-gram method (Continuous skip-gram model) and Continuous bag-of-words model (CBOW); both are shallow layers that are comprised of an input layer, a hidden layer (HL), and an output layer. The skip-gram and CBOW techniques utilize HL weight, which is attained throughout training to signify word vectors. While their intentions are the same, they vary in training models. The CBOW model forecasts the present word depending on its framework, utilizing many context words to predict the centre word. This methodology is efficient on tiny databases, and training is fast due to its use of context word data distribution. On the contrary, the skip-gram method forecasts the nearby context depending upon an assumed word, utilizing the centre word as an input to forecast its adjacent context word. However, skip-gram is best with larger databases and singular phrases, and its training advances at a moderately slower rate when compared to CBOW. The CBOW method concentrates on expecting a target word depending upon its surrounding context words.

The stages included in the CBOW technique are given below:

- Initialize: Context words were denoted utilizing one-hot encoded, creating an input layer where the dimension of

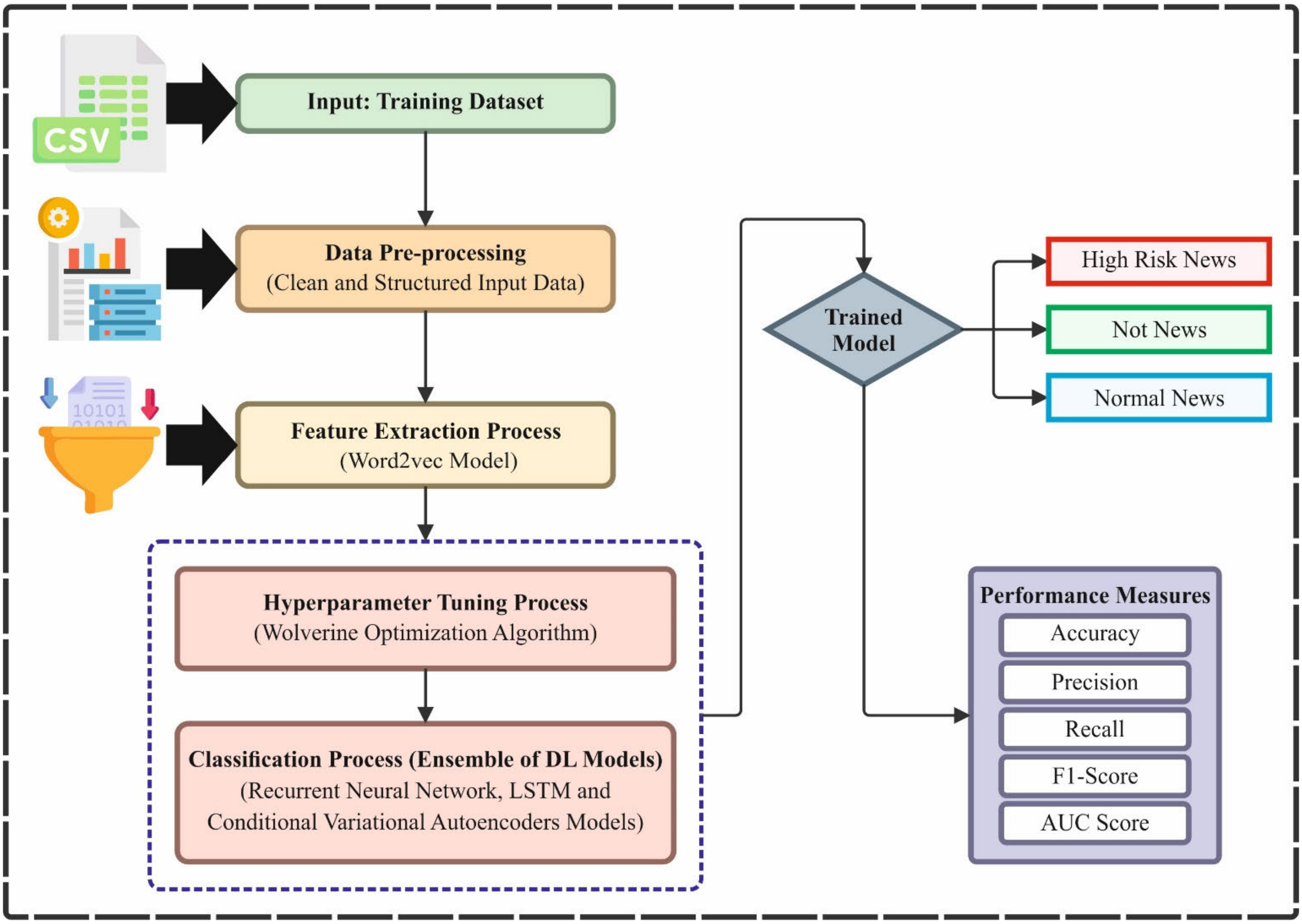
CSV
Input: Training Dataset
Data Pre-processing
(Clean and Structured Input Data)
10101
Feature Extraction Process
(Word2vec Model)
Hyperparameter Tuning Process
(Wolverine Optimization Algorithm)
Classification Process (Ensemble of DL Models)
(Recurrent Neural Network, LSTM and
Conditional Variational Autoencoders Models)
Trained
Model
High Risk News
Not News
Normal News
Performance Measures
Accuracy
Precision
Recall
F1-Score
AUC Score

FIGURE 2. The overall flow of the ETIC-EDLMOA model.

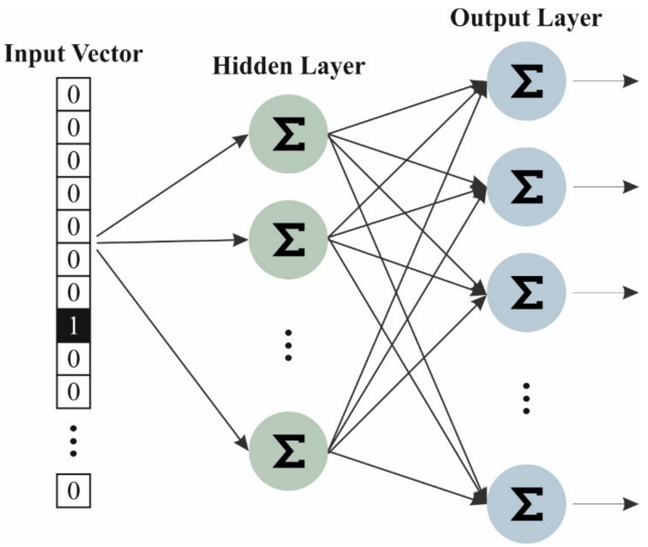

FIGURE 3. Structure of Word2vec model.

vocabulary is signified as V, and the amount of context words is denoted as C.

- HL Processing: Every one-hot encoding vector was multiplied by weight matrix W that links an input layer to the HL.
- Vector Averaging: The vectors attained from step 2 are averaged to create one vector at the HL.
- Processing of output Layer: This vector has been multiplied with additional weight matrix W', transitioning from HL to an output layer.

Prediction & Softmax Activation: The output layer experiences a softmax activation, a likelihood spread over the V-dimensional vocabulary. The most excellent probability index has been selected as the forecasted target word.

$$h = W^T x \tag{1}$$

$$u_j = W^{'T} h \tag{2}$$

$$y = \frac{\exp(\mathrm{u_j})}{\sum_{j'=1}^{V} \exp\left(u_{j'}\right)} \tag{3}$$

Here, h represents a HL, and W^T denotes a reverse of the weight of a vector (x). u_j refers to an output line to HL. The softmax activation function (y_i) utilizes an output value of the present line (u_j) as per the number of vocabulary words (V).

C. STAGE III: ENSEMBLE CLASSIFICATION PROCESS

The ensemble of DL models is employed for the classification process, comprising the RNN method, LSTM model, and CVAE technique. The ensemble models are chosen because they can handle intrinsic patterns in sequential data and generate robust representations. RNNs effectively capture temporal dependencies, making them ideal for sequential data. At the same time, LSTMs are an extension of RNNs that address long-term dependency issues, which are significant for tasks involving long sequences. CVAE, on the contrary, enables the modelling of data distributions in a probabilistic framework, improving the flexibility and generation of diverse feature representations. The ensemble approach utilizes the merits of each model: RNNs and LSTMs capture temporal and sequential dependencies, while CVAE assists in data generation and representation learning. This hybrid method increases model accuracy, reduces overfitting, and improves generalization, mainly when dealing with diverse and complex datasets, giving superior performance over individual models.

1) RNN CLASSIFIER

RNNs are a neural network class for handling sequential data, making them highly efficient for speech recognition, time series, financial predicting tasks, and natural language processing (NLP) [31]. Unlike conventional feed-forward NNs, which handle inputs autonomously, RNN features permit them to keep an HL, gradually taking the contextual and sequential connections between data points. This novel ability allows RNNs to handle data using temporal dependency, making them valuable for tasks while context is essential.

In Mathematics, the HL at tth time signified as h_t, refers to the task of either the present input x_t or the HL from the earlier time step h_{t-1}. This recursive procedure is stated as shown:

$$h_t = \sigma\left(W_h x_t + U_h h_{t-1} + b_h\right) \tag{4}$$

Meanwhile, W_h and U_h represent the weight matrix used for the present input x_t, and previous HL h_{t-1}, and b_h means biased term. σ denotes a sigmoid function, usually a sigmoid or $tanh$, which presents nonlinearity to the system. Nevertheless, RNNs face significant challenges like the exploding gradient and vanishing issues, which occur during the backpropagation through time (BPTT). These problems arise after the gradients grow and shrink mathematically as they are propagated over various layers, making it problematic for the method for learning longer-term dependency in the data.

2) LSTM CLASSIFIER

The LSTM structure has been specially intended to overwhelm the vanishing gradient problem, making it highly efficient at seizing longer-range dependency in sequential data [32]. The system characteristic of LSTM is a more composite cell structure that consists of memory cells and numerous gating methods for regulating the information flow, allowing the system to keep related information over extended periods while forgetting unrelated details. The main mechanisms of an LSTM cell are forget gate (f_t) and determining which data from the preceding state of the cell must be rejected. Input gate (i_t) defines novel information as something that must be deposited within the present state of the cell. Output gate (0_t) controls the output from the present state of the cell to the HL. Fig. 4 illustrates the LSTM classifier.

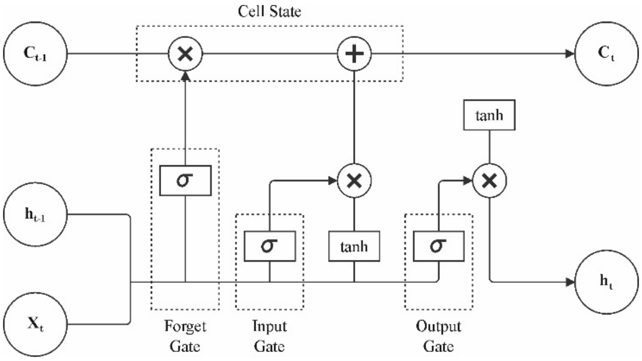

FIGURE 4. Architecture of CVAE.

The processes in the LSTM cell are mathematically described as demonstrated:

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{5}$$
$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \tag{6}$$
$$\tilde{C}_t = \tanh\left(W_c x_t + U_c h_{t-1} + b_c\right) \tag{7}$$
$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{8}$$
$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right) \tag{9}$$
$$h_t = o_t \odot \tanh\left(C_t\right) \tag{10}$$

On the other hand, C_t represents the state of the cell at tth time, and h_t refers to HL. The gating mechanisms are influenced by bias terms b and weight matrices W and U, whereas σ signifies the sigmoid function and $\odot$ characterizes element-to-element multiplication. The structure allows LSTMs to learn and keep longer-term dependency, making them appropriate for composite time-series predicting tasks using more extensive databases. Nevertheless, LSTMs are computationally expensive because of the numerous gates and memory cells.

3) CVAE CLASSIFIER

In recent decades, the AE concept has been the main element of emerging NNs [33]. The major AE application is to attain non-linear dimensionality, reducing massive datasets. But, thanks to current developments in hidden variable methods, they become prominent in generating application models. AE has dual main components, which are a decoder and an encoder. The mapping encoder inputs data (x) into a reduced dimension space encoded by generating a latent code (z). Afterwards, the decoding module tries to rebuild the approximate input data (x') using z. AE optimization usually focuses on reducing the reconstructed errors between the output and input data x' and x, respectively, by utilizing the backpropagation method with a different stochastic gradient descent.

VAE are reproductive methods that utilize DNNs to forecast variational distribution parameters. It has been shown that demonstration abilities have significantly enhanced those classical AE methods. The aim of trained VAE is attained across a variational Bayesian method for the true posterior distribution. The latent variable's $p_\theta(z|x)$ is expected to be inflexible due to the absence of closed-form limitation in the latent space. Bayes's theorem represents it.

$$p_\theta\left(z\,|\,x\right) = \frac{p_\theta\left(x\,|\,z\right)p_\theta\left(z\right)}{p_\theta\left(x\right)}, \tag{11}$$

The marginal likelihood $p_\theta(x)$ reading

$$p_\theta\left(x\right) = \int p_\theta\left(x\,|\,z\right)p_\theta\left(z\right)dz, \tag{12}$$

while $p_\theta(x|z)$ is the data utilized to implement $p_\theta(z)$ as the previous latent distribution. The VAE method must employ a distribution model, termed the variational inference method $q_\varphi(z|x)$, to approximate $p_\theta(z|x)$. To estimate parameterized distribution by encoding multi-layer perceptron with internal weights φ. In addition, $p_\theta(x|z)$ is a likelihood decoding multi-layer perceptron parameterized by weights θ. This research focuses on mapping the input data into its fundamental propagative distribution parameters for data augmentation goals. Thus, the input data x is changed into vectors μ and σ, related to co-variance parameters and Gaussian density mean function. The reparameterization trick

$$z = \mu + \sigma \odot \epsilon, \tag{13}$$

While $\odot$ represents the element-to-element outcome, then utilized to absorb σ and μ in the multi-variate Gaussian condition, leave the stochastic assets of an auxiliary vector $\epsilon \sim \mathcal{N}(0, I)$, wherever I is the entire individuality. At last, the decoding rebuilds the latent vector z to x', approximating the AE standards. VAEs intend to balance the exchange between data restoration possibilities or loss and regularization evaluated by Kullback-Leibler divergence (D_{KL}) to absorb important latent characteristics and create new data efficiently.

CVAE improves the fundamental VAE structure by preparing the propagative method with auxiliary variables. Similarly to the inference method, CVAEs also apply an encoding system increased by the auxiliary variables, $q_\varphi(z|x, c)$, while c is the situation. Model one summarizes the presented CVAE method training procedure. The main goal of training the CVAE is to attain optimum conditional Evidence Lower Bound (ELBO), which includes increasing the possibilities of monitoring data by examining the latent variable inside the method. The ELBO is

$$\begin{aligned} &ELBO(x, z, c) \\ &= \mathcal{E}\left[\log p_\theta\left(x\,|\,z, c\right)\right] - D_{KL}(q_\varphi\left(z\,|\,x, c\right)||p_\theta(z, c)), \end{aligned} \tag{14}$$

The primary term is the logarithmic expectation of data monitoring, and the next term is the Kullback-Leibler difference between the encoding distribution and previous latency. The negative ELBO, as the CVAE loss function ($\mathcal{L}_{CVAE}$), will be reduced after simplification states are utilized.

$$\begin{aligned} &\mathcal{L}_{CVAE} \\ &\simeq -\frac{1}{2}\sum_{k=1}^{K}\left(1 + log\sigma_k^2 - \mu_k^2 - \sigma_k^2\right) + \frac{1}{J}\sum_{j=1}^{J}\left\|x_j - x_j'\right\|_2^2, \end{aligned} \tag{15}$$

While $K = 16$ is the dimensional latent space and $J = 173$ is the dimensional data. Fig. 5 represents the structure of CVAE.

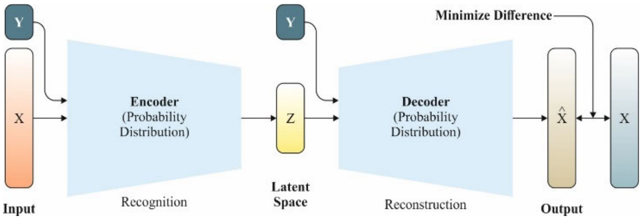

FIGURE 5. Architecture of CVAE.

D. STAGE IV: WOOA-BASED PARAMETER TUNING

Eventually, the hyperparameter fine-tuning process of the ensemble models is performed using the WoOA [34]. This metaheuristic optimization technique is appropriate for enhancing the performance of ML models by effectively tuning their hyperparameters. Unlike conventional methods such as grid or random search, which can be computationally expensive and time-consuming, the WoOA presents a more effective way to explore the parameter space. WoOA replicates the hunting behaviour of whales, giving a global search capability that can avoid local optima, making it especially beneficial for complex, high-dimensional optimization problems. Its ability to balance exploration and exploitation results in improved convergence rate and model performance. The flexibility and robustness of the WoOA model make it ideal for fine-tuning various models in dynamic environments, where hyperparameter optimization is crucial for achieving optimal results. Fig. 6 demonstrates the WoOA methodology.

The most significant land-dwelling sort of wolverine is mainly observed in remote alpine and subarctic tundra regions in the Northern Hemisphere and the Northern boreal forests. Between the behaviours of wolverines naturally and in the wild, the feeding tactics of this animal are significantly more prominent. Wolverine is consumed through dual methods, such as scavenging and hunting. In the scavenging tactic, the wolverine discovers the carrion and consumes it in the tracks of other hunters. In the hunting tactic, the wolverine attacks the live target, and after a fight-and-chase procedure, it destroys the prey and serves on it. These dual wolverine

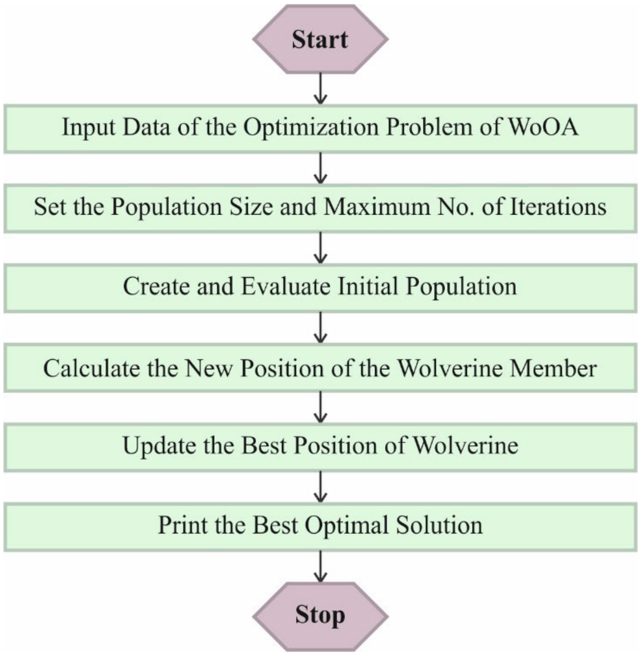
Start
Input Data of the Optimization Problem of WoOA
Set the Population Size and Maximum No. of Iterations
Create and Evaluate Initial Population
Calculate the New Position of the Wolverine Member
Update the Best Position of Wolverine
Print the Best Optimal Solution
Stop

FIGURE 6. **Architecture of WoOA.**

feeding strategies are intelligent processes whose mathematical modelling is employed to design the proposed WoOA approach, which is discussed below.

The WoOA model's design includes upgrading the location of the population associated with the problem-solving space by pretending the wolverine feeding behaviour. The wolverine contains dual approaches for feeding, such as hunting and scavenging. In the scavenging tactic, wolverines serve on empty carrion by moving beside the route of other hunters, which drop the remains of their destructions. In the hunting tactic, the wolverine first assaults the target, and after going over a fighting process, it destroys the prey and serves on it. In WoOA, it is said that in every iteration, every wolverine selects one at random of these two tactics with equivalent likelihood, and depending upon the simulation of the chosen tactic, its location in the problem-solving space is upgraded. Utilizing the Wolverine models is a decision-making procedure to define whether to search for food. In each iteration, every wolverine's location is upgraded depending on the 1st and 2nd tactics.

Upgrade procedure for ith wolverine

$$X_i\colon \begin{cases} \textit{based on scavenging strategy}, & r_p \leq 0.5 \\ \textit{based on hunting strategy}, & \textit{else} \end{cases} \tag{16}$$

where r_p denotes a number within the range of [0, 1].

Strategy 1: Searching Strategy (Exploration Stage)

The WoOA strategy involves upgrading the position of the population member in the problem-solving space over modelling the feeding behaviour of wolverines on meat. In this approach, the wolverine, to obtain the carrion, obeys the route of other hunters who have left the rest of their kills. Pretending the wolverine's drive in response to hunters' outcomes is a very dissimilar technique to explore the solution space. This model presents a significant adaptation in the locations of the population members, efficiently modifying their tracks in the problem-solving regions. By accepting this tactic, the technique attains a more complete search of probable solutions. These massive positional modifications extend the searching region and improve the model's capability to uncover different and optimum solutions. Thus, the technique enhances the ability of the model to execute a complete global search, assisting it in discovering intricate problem spaces and finding a new solution. The scavenging tactic contains an exploration stage that depends upon pretending the wolverine's drive to get a carrion.

In the design of WoOA, the position of other members in a superior objective value of the function is measured as the predators' location who are directing to relieve the rest of their destroys for every wolverine, as stated in

$$\begin{aligned} CP_i = \{X_k\colon F_k < F_i \textit{ and } k \neq i, \} \textit{ where} \\ i = 1, 2, \ldots, N \textit{ and } k \in \{1, 2, \ldots, N\} \end{aligned} \tag{17}$$

While CP_i denotes a group of candidate predators' positions for ith Wolverine, X_k refers to a population member with a well-objective function value compared to *ith* Wolverine, and F_k represents an objective function value.

The WoOA design assumes that the wolverine randomly selects a hunter's location from the CP_j group and travels near it next. As the wolverine travels near the chosen predator to extend the carrion, a novel proposed location for the particular member was calculated. If the objective value of the function is improved, this developed location replaces the aforementioned position of the subsequent member as per the below-mentioned equations:

$$x_{i,j}^{S1} = x_{i,j} + r_{i,j} \cdot \left(SP_{i,j} - I_{i,j} \cdot x_{i,j}\right), \tag{18}$$

$$X_i = \begin{cases} X_i^{S1}, & F_i^{S1} \leq F_i, \\ X_i, & \textit{else}, \end{cases} \tag{19}$$

Here, SP_i denotes the selected predator for ith wolverine, $SP_{i,j}$ indicates the jth dimension of the predator. X_i^{S1} is the newly calculated location of ith wolverine, depending upon the searching tactic used in the developed WoOA. In the same way, $x_{i,j}^{S1}$ refers to a jth dimension of this novel location. F_i^{S1} signifies the objective function value linked with the ith wolverine at its novel location. Furthermore, $r_{i,j}$ is a randomly generated number equally within the range of [0, 1], and $I_{i,j}$ is a dual randomly produced number 1 and 2.

Strategy 2: Hunting Strategy (Exploration and Exploitation Stages)

The second tactic of WoOA involves upgrading the location of the population members in the problem-solving space by pretending to be the hunting behaviour of a wolverine. This procedure reflects the wolverine's conventional technique for searching, where it initially assaults live prey and is then involved in a chase and fight before securing and feeding on its kill. The population members' upgrade depends upon this hunting approach, which contains dual stages such as exploration over pretending the wolverine's drive near the prey and exploitation by modelling the chase and fight procedure among the wolverine and its prey.

Stage 1: Attack (exploration phase)

In this phase of WoOA, the population's position in the problem-solving space is altered as an outcome of suggesting the procedure of a wolverine attacking its prey. By emulating the wolverine's movements throughout a search, significant alterations are presented to the population's location, thus improving the exploration abilities of WoOA in directing over the problem-solving region. While designing the WoOA, the location of the finest population member is compared to the prey's position. A novel potential location is computed by pretending the wolverine's tactic is near the chosen member (the "prey") for every member of WoOA. Then, if the objective function determines an upgrade, this newly developed location substitutes the earlier one for the equivalent member, as explained below:

$$x_{i,j}^{P1} = x_{i,j} + r_{i,j}\cdot \left(Prey_j - I_{i,j}\cdot x_{i,j}\right), \tag{20}$$

$$X_i = \begin{cases} X_i^{P1}, & F_i^{P1} \leq F_i, \\ X_i, & else, \end{cases} \tag{21}$$

Here, *Prey* represents the finest population member as prey, $Prey_j$ refers to *the jth* dimension, and X_i^{P1} indicates the newly calculated location for *ith* wolverine, defined by the initial stage of the hunting approach in the developed WoOA. At the same time, $x_{i,j}^{P1}$ denotes the *jth* dimension of this novel location. F_i^{P1} refers to an objective function value of *ith* wolverine. Furthermore, $r_{i,j}$ refers to randomly generated values within the range of [0, 1] and $I_{i,j}$ is dual numbers randomly allocated as both 1 and 2.

Phase 2: Fighting and chasing (exploitation stage)

Throughout this WoOA stage, the population members' problem-solving positions are altered by modelling the search and hunt connections between the wolverine and its prey. By simulating the movements of the wolverine during chasing, small changes are proposed to the locations of the population member, thus improving the WoOA's ability to do local searching in the problem-solving space. The model of WoOA affects these contacts near the chasing position. Through modelling the wolverine's activities throughout the pursuit and hunt, a novel location is computed for every WoOA member. If this original location enhances an objective function value, it switches the member's preceding spot.

$$x_{i,j}^{P2} = x_{i,j} + \left(1 - 2r_{i,j}\right)\cdot \frac{ub_j - lb_j}{t} \tag{22}$$

$$X_i = \begin{cases} X_i^{P2}, & F_i^{P2} \leq F_i \\ X_i, & else \end{cases} \tag{23}$$

Here, X_i^{P2} represents the novel location defined for *ith* wolverine depending upon the 2nd phase of hunting tactic used in the developed WoOA. $x_{i,j}^{P2}$ signifies the *jth* dimension of this upgraded location. F_i^{P2} signifies the objective function value linked with the *ith* wolverine at its new spot. Furthermore, $r_{i\ and\ j}$ are randomly produced numbers within the range of [0, 1], and t signifies the iteration count.

The WoOA arises a fitness function (FF) to realize the enhanced performance of the classifier. It determines a positive numeral for denoting an improved efficiency of the candidate solution. Here, the minimization of the classifier rate of error is reflected as FF. Its mathematical calculation is expressed below in Eq. (24).

$$\begin{aligned} fitness\left(x_i\right) &= Classifier\ Error\ Rate\left(x_i\right) \\ &= \frac{No.\ of\ misclassified\ samples}{Total\ no.\ of\ samples} \times 100 \end{aligned} \tag{24}$$

IV. EXPERIMENTAL VALIDATION

The performance evaluation of the ETIC-EDLMOA methodology is studied using the CybAttT dataset [35]. The dataset contains 8800 tweets under three classes, as illustrated in Table 2. Table 3 signifies the sample texts.

TABLE 2. Details of dataset.

Labels	No. Tweets
"High Risk News"	800
"Normal News"	3000
"Not News"	5000
Total Tweets	8800

TABLE 3. Sample texts.

S.no	Labels	Text
1	High-Risk News	"Ransomware sent North Carolina A&amp;T University scrambling to restore services | Ars Technica https://t.co/2PVXsbz00U"
2	High-Risk News	"OpenSea Reports Stolen Email Addresses in Data Breach, Warns Users About Phishing Possibility https://t.co/XQIqyete5W via @coinjupiter"
3	Normal News	"Snap-on Tools Hit by Cyberattack Claimed by Conti Ransomware Gang"
4	Normal News	"LockBit ransomware gang lurked in a U.S. gov network for months."
5	Not News	"Microsoft Details New Security Features for Windows 11"
6	Not News	"Software-as-a-Service Rules the Cloud"

Fig. 7 establishes the confusion matrix the ETIC-EDLMOA methodology produces under various epochs. The outcomes indicate that the ETIC-EDLMOA methodology effectively detects and recognizes all class labels precisely.

The attack detection of the ETIC-EDLMOA approach is determined under distinct epochs in Table 4 and Fig. 8. The table values state that the ETIC-EDLMOA approach correctly recognized all the samples. On 500 epochs, the ETIC-EDLMOA approach provides an average $accu_y$ of 97.61%, $prec_n$ of 94.77%, $reca_l$ of 93.67%, $F1_{score}$ of 94.18%, and AUC_{score} of 95.83%. Moreover, dependent on 1000 epochs, the ETIC-EDLMOA method gives an average $accu_y$ of 98.31%, $prec_n$ of 96.83%, $reca_l$ of 95.01%, $F1_{score}$ of 95.86%, and AUC_{score} of 96.76%. Also, based on

2000 epochs, the ETIC-EDLMOA method offers an average $accu_y$ of 98.24%, $prec_n$ of 96.44%, $reca_l$ of 94.95%, $F1_{score}$ of 95.65%, and AUC_{score} of 96.70%. In addition, with 2500 epochs, the ETIC-EDLMOA model delivers an average $accu_y$ of 98.51%, $prec_n$ of 96.51%, $reca_l$ of 96.23%, $F1_{score}$ of 96.36%, and AUC_{score} of 97.48%. Lastly, concerning 3000 epochs, the ETIC-EDLMOA model provides an average $accu_y$ of 98.30%, $prec_n$ of 97.01%, $reca_l$ of 94.65%, $F1_{score}$ of 95.75%, and AUC_{score} of 96.56%.

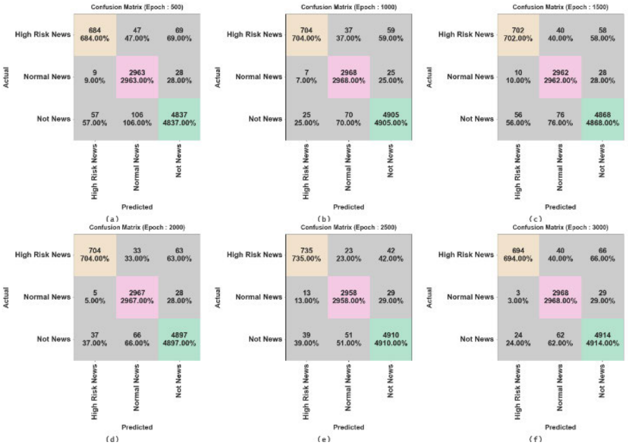

FIGURE 7. Confusion matrix of ETIC-EDLMOA methodology (a-f) Epochs 500-3000.

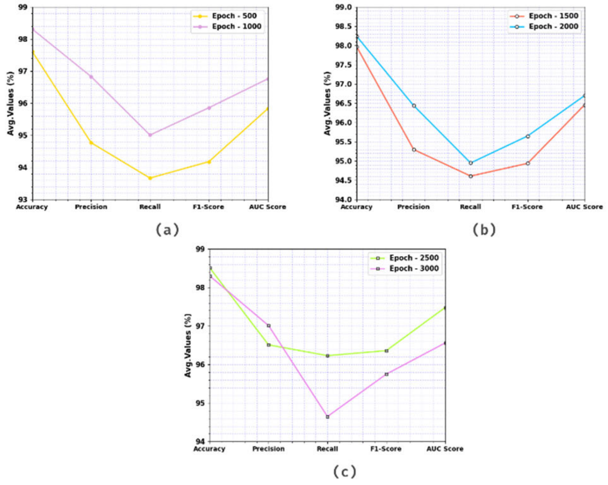

FIGURE 8. Average of ETIC-EDLMOA technique (a) Epochs 500-1000, (b) Epochs 1500-2000, and (c) Epochs 2500-3000.

Fig. 9 illustrates the training (TRA) $accu_y$ and validation (VAL) $accu_y$ analysis of the ETIC-EDLMOA model under different epochs. The $accu_y$ analysis is calculated across the range of 0-3000 epochs. The figure highlighted that the TRA and VAL $accu_y$ analysis demonstrated an increasing tendency, which informed the capacity of the ETIC-EDLMOA approach with superior outcomes over multiple iterations. At the same time, the TRA and VAL $accu_y$ leftovers closer across the epochs, which specifies inferior overfitting and demonstrates higher outcomes of the ETIC-EDLMOA approach, guaranteeing dependable prediction on unseen samples.

TABLE 4. Attack detection of ETIC-EDLMOA model under various epochs.

Classes	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{score}$	AUC_{score}
Epoch - 500					
High-Risk News	97.93	91.20	85.50	88.26	92.34
Normal News	97.84	95.09	98.77	96.89	98.06
Not News	97.05	98.03	96.74	97.38	97.09
Average	97.61	94.77	93.67	94.18	95.83
Epoch - 1000					
High-Risk News	98.55	95.65	88.00	91.67	93.80
Normal News	98.42	96.52	98.93	97.71	98.54
Not News	97.97	98.32	98.10	98.21	97.94
Average	98.31	96.83	95.01	95.86	96.76
Epoch - 1500					
High-Risk News	98.14	91.41	87.75	89.54	93.46
Normal News	98.25	96.23	98.73	97.47	98.37
Not News	97.52	98.26	97.36	97.81	97.55
Average	97.97	95.30	94.61	94.94	96.46
Epoch - 2000					
High-Risk News	98.43	94.37	88.00	91.07	93.74
Normal News	98.50	96.77	98.90	97.82	98.60
Not News	97.80	98.18	97.94	98.06	97.77
Average	98.24	96.44	94.95	95.65	96.70
Epoch - 2500					
High-Risk News	98.67	93.39	91.87	92.63	95.61
Normal News	98.68	97.56	98.60	98.08	98.66
Not News	98.17	98.57	98.20	98.39	98.17
Average	98.51	96.51	96.23	96.36	97.48
Epoch - 3000					
High-Risk News	98.49	96.26	86.75	91.26	93.21
Normal News	98.48	96.68	98.93	97.79	98.59
Not News	97.94	98.10	98.28	98.19	97.89
Average	98.30	97.01	94.65	95.75	96.56

Fig. 10 exposes the TRA loss (TRALOS) and VAL loss (VALLOS) curves of the ETIC-EDLMOA technique under dissimilar epochs. The loss values are computed within the

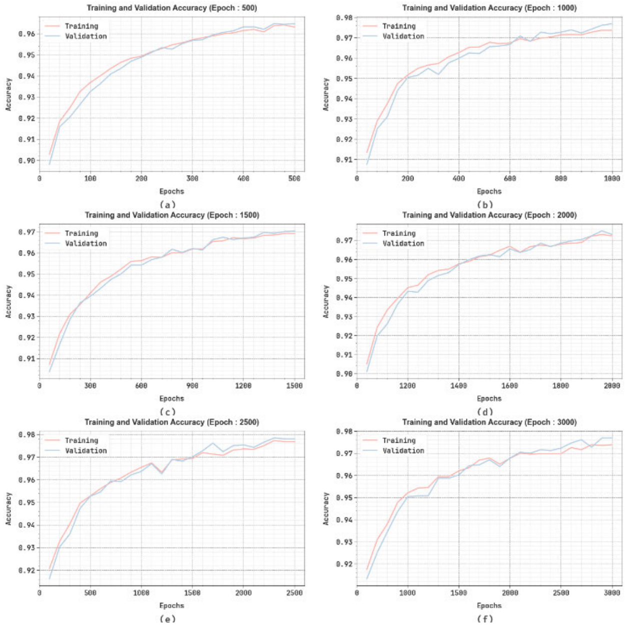

FIGURE 9. *$Accu_y$* **curve of ETIC-EDLMOA model (a-f) Epochs 500-3000.**

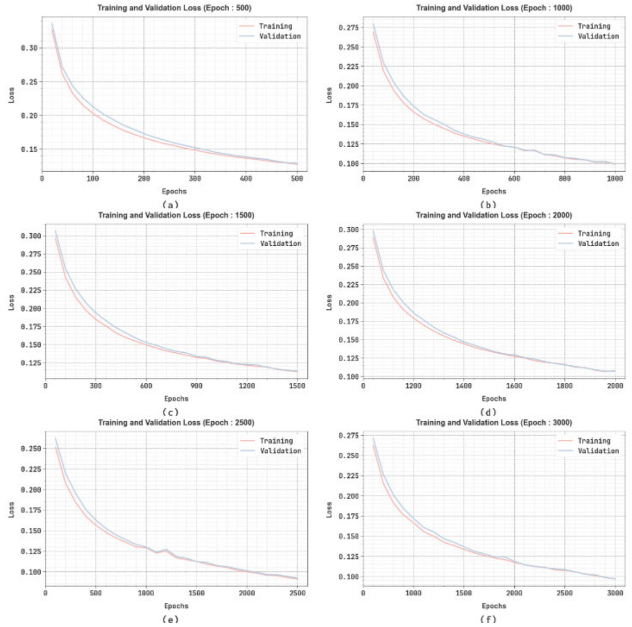

FIGURE 10. **Loss analysis of ETIC-EDLMOA model (a-f) Epochs 500-3000.**

range of 0-3000 epochs. It is denoted that the TRALOS and VALLOS values exemplify a diminishing trend, reporting the ETIC-EDLMOA methodology's capacity to balance a trade-off between simplification and data fitting. The constant reduction in loss values guarantees the superior outcome of the ETIC-EDLMOA methodology and tunes the prediction results over time.

In Fig. 11, the precision-recall (PR) analysis of the ETIC-EDLMOA approach under different epochs clarifies its performance by plotting Precision beside Recall for three classes. The outcome illustrates that the ETIC-EDLMOA approach continuously accomplishes better PR values over other classes, signifying its capacity to conserve an essential portion of true positive predictions between all positive predictions (precision) while taking a massive proportion of actual positives (recall). The constant rise in PR analysis amongst three class labels shows the efficiency of the ETIC-EDLMOA technique in the classification procedure.

In Fig. 12, the ROC analysis of the ETIC-EDLMOA methodology under different epochs is examined. The results imply that the ETIC-EDLMOA approach achieves improved ROC analysis over every class, indicating an essential capacity for discerning the class labels. This consistent trend of higher ROC analysis over multiple class labels indicates the capable performance of the ETIC-EDLMOA approach in predicting classes, highlighting the robust nature of the classification procedure.

Table 5 and Fig. 13 study the comparison outcomes of the ETIC-EDLMOA methodology with existing techniques [19], [20], [29]. The simulation results stated that the ETIC-EDLMOA methodology outperformed superior performances. Based on $accu_y$, the ETIC-EDLMOA methodology has a higher $accu_y$ of 98.51%. In contrast, the CV-DT, CV-KNN, CV-MNB, TF-IDF-LR, TF-IDF-SVM, DistilBERT, and RoBERTa approaches have lesser $accu_y$ of 95.17%, 95.07%, 95.37%, 96.45%, 96.92%, 97.26%, and 97.24%, respectively. At the same Time, based on $accu_y$, the ETIC-EDLMOA methodology has a better $Prec_n$ of 96.51%, where the CV-DT, CV-KNN, CV-MNB, TF-IDF-LR, TF-IDF-SVM, DistilBERT, and RoBERTa models have lower $Prec_n$ of 84.26%, 89.27%, 86.56%, 91.96%, 92.17%, 96.02%, and 96.17%, correspondingly. Simultaneously, dependent on $F1_{score}$, the ETIC-EDLMOA methodology has a maximum $F1_{score}$ of 96.36%. In contrast, the CV-DT, CV-KNN, CV-MNB, TF-IDF-LR, TF-IDF-SVM, DistilBERT, and RoBERTa models have minimal $F1_{score}$ of 82.75%, 82.36%, 82.55%, 85.16%, 86.46%, 96.02%, and 96.10%, correspondingly.

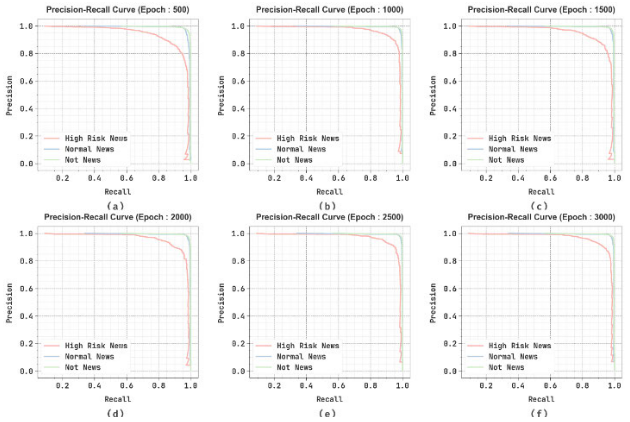

FIGURE 11. **PR curve of ETIC-EDLMOA model (a-f) Epochs 500-3000.**

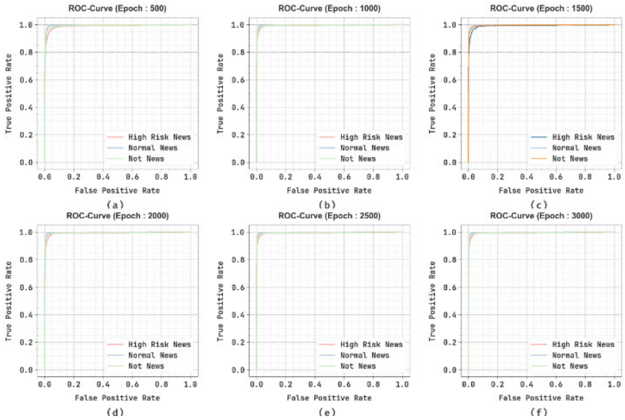

FIGURE 12. ROC analysis of ETIC-EDLMOA model (a-f) Epochs 500-3000.

TABLE 5. Comparative results of ETIC-EDLMOA technique with existing models.

Techniques	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{score}$
CV-DT	95.17	84.26	81.36	82.75
CV-KNN Method	95.07	89.27	77.26	82.36
CV-MNB Model	95.37	86.56	81.08	82.55
TF-IDF-LR Model	96.45	91.96	80.35	85.16
TF-IDF-SVM	96.92	92.17	82.36	86.46
DistilBERT Method	97.26	96.02	96.00	96.02
RoBERTa system	97.24	96.17	96.07	96.10
XAI	96.01	89.77	87.07	88.16
SHAP	96.18	94.79	82.89	88.03
GNB	96.15	92.36	86.63	88.03
ETIC-EDLMOA	98.51	96.51	96.23	96.36

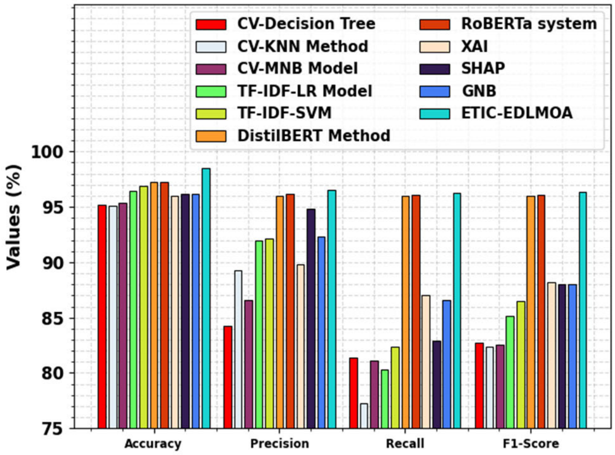

FIGURE 13. Comparative analysis of ETIC-EDLMOA technique with existing models.

Table 6 and Fig. 14 illustrate the computational time (CT) analysis of the ETIC-EDLMOA methodology with existing techniques.

TABLE 6. CT analysis of ETIC-EDLMOA technique with existing models.

Techniques	CT (sec)
CV-DT	15.15
CV-KNN Method	23.44
CV-MNB Model	22.02
TF-IDF-LR Model	14.03
TF-IDF-SVM	24.31
DistilBERT Method	18.85
RoBERTa system	19.05
XAI	13.81
SHAP	21.68
GNB	10.86
ETIC-EDLMOA	8.10

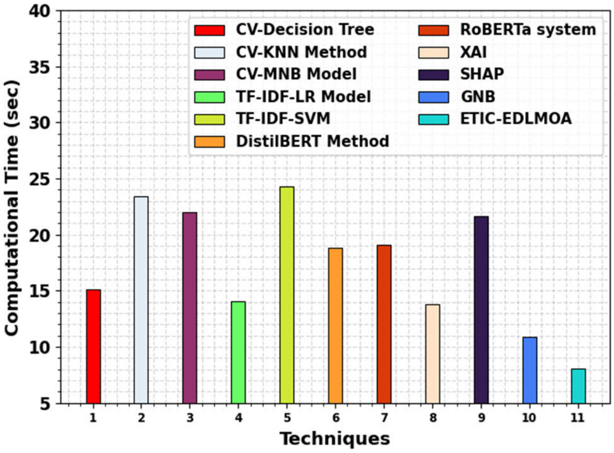

FIGURE 14. CT analysis of ETIC-EDLMOA technique with existing models.

V. CONCLUSION

In this study, the ETIC-EDLMOA model is proposed. The presented ETIC-EDLMOA model's main aim is to detect and mitigate network attacks in cybersecurity effectively. To accomplish that, the ETIC-EDLMOA model contains distinct stages such as data pre-processing, feature extraction, ensemble classification process, and parameter optimizer. Initially, the ETIC-EDLMOA model undergoes a data pre-processing stage to ensure clean and structured input data for analysis. Besides, the Word2vec model has been deployed for the extraction of features. The ensemble of DL models was employed for the classification process, including the RNN method, LSTM model, and CVAE technique. Eventually, the hyperparameter fine-tuning process of the ensemble models is performed using the WoOA. A comprehensive range of simulation analyses is conducted to ensure the improved performance of the ETIC-EDLMOA method on the CybAttT dataset. The comparison study of the ETIC-EDLMOA method illustrated a superior accuracy value of 98.51% over existing techniques.

DATA AVAILABILITY STATEMENT

The data supporting this study's findings are openly available at https://github.com/HudaLughbi/CybAttT, reference number [35].

REFERENCES

[1] P. Ranade, A. Piplai, S. Mittal, A. Joshi, and T. Finin, "Generating fake cyber threat intelligence using transformer-based models," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2021, pp. 1–9.

[2] D. Schlette, F. Böhm, M. Caselli, and G. Pernul, "Measuring and visualizing cyber threat intelligence quality," *Int. J. Inf. Secur.*, vol. 20, no. 1, pp. 21–38, Feb. 2021.

[3] T. Schaberreiter, V. Kupfersberger, K. Rantos, A. Spyros, A. Papanikolaou, C. Ilioudis, and G. Quirchmayr, "A quantitative evaluation of trust in the quality of cyber threat intelligence sources," in *Proc. 14th Int. Conf. Availability, Rel. Secur.*, Aug. 2019, pp. 1–10.

[4] J. Robertson, A. Diab, E. Marin, E. Nunes, V. Paliath J. Shakarian, and P. Shakarian, *Darkweb Cyber Threat Intelligence Mining*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[5] V. G. Li, M. Dunn, P. Pearce, D. McCoy, G. M. Voelker, S. Savage, and K. Levchenko, "Reading the tea leaves: A comparative analysis of threat intelligence," in *Proc. 28th USENIX Secur. Symp.*, 2019, pp. 851–867.

[6] H. Griffioen, T. Booij, and C. Doerr, "Quality evaluation of cyber threat intelligence feeds," in *Proc. 18th Int. Conf. Appl. Cryptogr. Netw. Secur. (ACNS)*, Rome, Italy, 2020, pp. 277–296.

[7] P. Gao, F. Shao, X. Liu, X. Xiao, Z. Qin, F. Xu, P. Mittal, S. R. Kulkarni, and D. Song, "Enabling efficient cyber threat hunting with cyber threat intelligence," in *Proc. IEEE 37th Int. Conf. Data Eng. (ICDE)*, Apr. 2021, pp. 193–204.

[8] W. Yang and K.-Y. Lam, "Automated cyber threat intelligence reports classification for early warning of cyber attacks in next generation SOC," in *Proc. Int. Conf. Inf. Commun. Secur.*, 2019, pp. 145–164.

[9] Y. Gao, X. Li, H. Peng, B. Fang, and P. S. Yu, "HinCTI: A cyber threat intelligence modeling and identification system based on heterogeneous information network," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 2, pp. 708–722, Feb. 2022.

[10] A. Sleem and I. Elhenawy, "Enhancing cyber threat intelligence sharing through a privacy-preserving federated learning approach," *J. Cybersecur. Inf. Manage.*, vol. 9, no. 2, pp. 51–59, 2022.

[11] T. Chen, H. Zeng, M. Lv, and T. Zhu, "CTIMD: Cyber threat intelligence enhanced malware detection using API call sequences with parameters," *Comput. Secur.*, vol. 136, Jan. 2024, Art. no. 103518.

[12] Z. Yu, J. Wang, B. Tang, and L. Lu, "Tactics and techniques classification in cyber threat intelligence," *Comput. J.*, vol. 66, no. 8, pp. 1870–1881, Aug. 2023.

[13] K. Ahmed, S. K. Khurshid, and S. Hina, "CyberEntRel: Joint extraction of cyber entities and relations using deep learning," *Comput. Secur.*, vol. 136, Jan. 2024, Art. no. 103579.

[14] F. Sufi, "A global cyber-threat intelligence system with artificial intelligence and convolutional neural network," *Decis. Analytics J.*, vol. 9, Dec. 2023, Art. no. 100364.

[15] A. T. Azar, S. U. Amin, M. A. Majeed, A. Al-Khayyat, and I. Kasim, "Cloud-cyber physical systems: Enhanced metaheuristics with hierarchical deep learning-based cyberattack detection," *Eng., Technol. Appl. Sci. Res.*, vol. 14, no. 6, pp. 17572–17583, Dec. 2024.

[16] G. Bhandari, A. Lyth, A. Shalaginov, and T.-M. Grønli, "Distributed deep neural-network-based middleware for cyber-attacks detection in smart IoT ecosystem: A novel framework and performance evaluation approach," *Electronics*, vol. 12, no. 2, p. 298, Jan. 2023.

[17] Y. Zhang, J. Chen, Z. Cheng, X. Shen, J. Qin, Y. Han, and Y. Lu, "Edge propagation for link prediction in requirement-cyber threat intelligence knowledge graph," *Inf. Sci.*, vol. 653, Jan. 2024, Art. no. 119770.

[18] C.-M. Chen, F.-H. Hsu, and J.-N. Hwang, "Useful cyber threat intelligence relation retrieval using transfer learning," in *Proc. Eur. Interdiscipl. Cybersecur. Conf.*, Jun. 2023, pp. 42–46.

[19] A. A. Mamun, H. Al-Sahaf, I. Welch, and S. Camtepe, "Genetic programming for enhanced detection of advanced persistent threats through feature construction," *Comput. Secur.*, vol. 149, Feb. 2025, Art. no. 104185.

[20] S. A. Sharaf, M. Ragab, N. Albogami, A. Al-Malaise Al-Ghamdi, M. F. Sabir, L. A. Maghrabi, E. B. Ashary, and H. Alaidaros, "Advanced mathematical modeling of mitigating security threats in smart grids through deep ensemble model," *Sci. Rep.*, vol. 14, no. 1, p. 23069, Oct. 2024.

[21] C. S. Kalutharage, X. Liu, and C. Chrysoulas, "Neurosymbolic learning and domain knowledge-driven explainable AI for enhanced IoT network attack detection and response," *Comput. Secur.*, vol. 151, Apr. 2025, Art. no. 104318.

[22] A. Baluguri, V. Pasumarthy, I. Roy, B. Gupta, and N. Rahimi, "Optimizing network security via ensemble learning: A Nexus with intrusion detection," *J. Inf. Secur.*, vol. 15, no. 4, pp. 545–556, 2024.

[23] F. S. Alrayes, M. Maray, A. Alshuhail, K. M. Almustafa, A. A. Darem, A. M. Al-Sharafi, and S. D. Alotaibi, "Privacy-preserving approach for IoT networks using statistical learning with optimization algorithm on high-dimensional big data environment," *Sci. Rep.*, vol. 15, no. 1, p. 3338, Jan. 2025.

[24] R. K. Batchu, T. Bikku, S. Thota, H. Seetha, and A. A. Ayoade, "A novel optimization-driven deep learning framework for the detection of DDoS attacks," *Sci. Rep.*, vol. 14, no. 1, p. 28024, Nov. 2024.

[25] M. Antonijevic, M. Zivkovic, M. Djuric Jovicic, B. Nikolic, J. Perisic, M. Milovanovic, L. Jovanovic, M. Abdel-Salam, and N. Bacanin, "Intrusion detection in metaverse environment Internet of Things systems by metaheuristics tuned two level framework," *Sci. Rep.*, vol. 15, no. 1, p. 3555, Jan. 2025.

[26] D. Kumari, A. Sinha, S. Dutta, and P. Pranav, "Optimizing neural networks using spider monkey optimization algorithm for intrusion detection system," *Sci. Rep.*, vol. 14, no. 1, p. 17196, 2024.

[27] R. R. Papalkar and A. S. Alvi, "Enhancing IoT security: A creative swagger optimization algorithm for DDoS defence," *Netw., Comput. Neural Syst.*, vol. 2025, pp. 1–39, Jan. 2025.

[28] T. S. Oyinloye, M. O. Arowolo, and R. Prasad, "Enhancing cyber threat detection with an improved artificial neural network model," *Data Sci. Manage.*, vol. 8, no. 1, pp. 107–115, Mar. 2025.

[29] H. Lughbi, M. Mars, and K. Almotairi, "CybAttT: A dataset of cyberattack news tweets for enhanced threat intelligence," *Data*, vol. 9, no. 3, p. 39, Feb. 2024.

[30] P. Rakshit and A. Sarkar, "A supervised deep learning-based sentiment analysis by the implementation of word2 Vec and GloVe embedding techniques," *Multimedia Tools Appl.*, vol. 84, no. 2, pp. 979–1012, Apr. 2024.

[31] A. Praveenkumar, G. K. Jha, S. D. Madival, A. Lama, and R. R. Kumar, "Deep learning approaches for potato price forecasting: Comparative analysis of LSTM, bi-LSTM, and AM-LSTM models," *Potato Res.*, vol. 2024, pp. 1–23, Oct. 2024.

[32] Y. Gao, N. Wang, and F. Li, "Steering drilling wellbore trajectory prediction based on the NOA-LSTM-FCNN method," *Sci. Rep.*, vol. 15, no. 1, p. 5215, Feb. 2024.

[33] L. M. A. Nagasingha, C. L. Bérubé, and C. J. M. Lawley, "A balanced mineral prospectivity model of Canadian magmatic Ni (±Cu±Co±PGE) sulphide mineral systems using conditional variational autoencoders," *Ore Geol. Rev.*, vol. 175, Dec. 2024, Art. no. 106329.

[34] T. Hamadneh, B. Batiha, O. Alsayyed, F. Werner, Z. Monrazeri, M. Dehghani, and K. Eguchi, "Using the novel wolverine optimization algorithm for solving engineering applications," *Comput. Model. Eng. Sci.*, vol. 141, no. 3, pp. 2253–2323, 2024.

[35] Accessed: Jul. 2, 2024. [Online]. Available: https://github.com/HudaLughbi/CybAttT

SRIPADA NSVSC RAMESH received the degree from Andhra University, in 1998, and the M.Tech. degree in IT from IASE University, in 2005. He is currently pursuing the Ph.D. degree in CSE with Arni University, Himachal Pradesh. He has been working with the Aditya College of Engineering and Technology (A), Surampalem, Andhra Pradesh, India, for the past 17 years. He has a total of more than 24 years of teaching experience. He is an Associate Professor of CSE and an Officer In Charge of Examinations with the Aditya College of Engineering and Technology (A). He published 12 journal articles, six conference papers, and two patents. His research interests include artificial intelligence, software cost estimation, machine learning, mobile computing, cloud computing, and soft computing. He received three awards from several organizations in recognition of his academic and research work, including one international award Innovative Technologist and Dedicated Teaching Professional Award from Malaysia. He organized and attended several workshops, FDPs, and conferences, and gave guest lectures at other colleges.

BADER MOHAMMED M. AL FARDAN received the B.Sc. degree from the Department of Industrial Engineering, King Khalid University (KKU), Saudi Arabia, in 2015, the M.Sc. degree from the Industrial Engineering and Operation Management Department, University of Central Florida (UCF), in 2021, and the Ph.D. degree in industrial engineering from UCF, in 2023. Currently, he is working with KKT as an Assistant Professor.

C. S. S. ANUPAMA received the Ph.D. degree in the area of wireless communications from Jawaharlal Nehru Technological University, Kakinada, India. She is currently an Associate Professor with the V. R. Siddhartha Engineering College, Vijayawada, India. She is actively engaged in research in artificial intelligence, machine learning, and deep learning domains related to medical and agricultural fields. She has 36 publications to her credit in various SCI, Scopus indexed journals, and international conferences. She was granted one international patent titled ''Wireless Networking of Medical Equipment's to Mobile Application for Paperless Clinic'' and published two national patents titled ''Deep Learning Based Intelligent Method for EEG Signal Analysis for Neurological Disorders Diagnosis'' and ''YOUPEP - Method and System for Building Network Between Users.'' She received a grant from the All India Council for Technical Education (AICTE), India, for conducting a short-term training program on ''Trends and Challenges in Medical Image Analysis Through Deep Learning Algorithms.'' She is executing two project grants titled ''System Diagnosis of Skin Cancer using Ensemble Machine Learning Models'' and ''Development of Health Monitoring System.'' She is a Life Member of the Biomedical Engineering Society of India (BMESI), the Indian Society for Technical Education (ISTE), and the Institution of Electronics and Telecommunication Engineers (IETE).

KOLLATI VIJAYA KUMAR received the Ph.D. degree in computer science and engineering from Karpagam University, Coimbatore. He has 18 years of teaching experience. He is currently an Associate Professor with the Department of CSE, GITAM School of Technology, GITAM University, Visakhapatnam, Andhra Pradesh, India. He has published 29 articles in various international and national journals. His research interests include wireless networking, big data, data analytics, network security, machine learning, cloud computing, and information security.

SEONGSOO CHO received the Bachelor of Science degree in advertising informatics from Gwangju University, in 1993, the Master of Science in publishing and magazine from Kyung Hee University, in 1996, and the Ph.D. degree in electronic engineering from Kwangwoon University, Seoul, South Korea, in 2010. He has held various Research Professor positions at esteemed institutions, such as Kwangwoon University, Mokpo National University, Soongsil University, and Chosun University. He is currently a Researcher with the Department of Convergence Science, Kongju National University, Gongju, South Korea. His dedication to academia and research underscores his commitment to advancing knowledge and innovation in the field of software convergence. In addition to his academic pursuits, he actively engages in professional and social activities, serving as the Chairperson of the Board of Directors for ICT-AES, Seoul, and contributing as an Editorial Board Member for prestigious journals, such as the *International Journal of Advanced Social Sciences* and the *International Journal of Advanced Engineering*.

SRIJANA ACHARYA received the B.S. degree in computer application from MCRP University, India, in 2003, and the M.S. degree in information and communication engineering and the Ph.D. degree in digital convergence business from Yeungnam University, in 2014 and 2021, respectively. She is currently a Postdoctoral Researcher with the Department of Convergence Science, Kongju National University, South Korea. Her research interests include webometrics, open data, data security, SNS security, SNS analysis, knowledge management, and digital convergence. She received various scholarships for the M.S. and Ph.D. studies.

CHEOLHEE YOON received the bachelor's degree in software engineering from Hansung University, in February 2004, the master's degree from the Digital Forensic Department, Korea University, in September 2016, and the Ph.D. degree in technology policy from Yonsei University, in February 2023. Since June 2017, he has been a Researcher with the Police Science Institute, Korean National Police University, Asan. His research interests include cutting-edge areas, such as autonomous cybersecurity, machine learning, and cybersecurity.