ISSN: 2093-5374 / E-ISSN: 2093-5382

Termite Apriori Prediction Framework for Detecting Cyber Threats in Social Networks

N. Sheba Pari[1], and Dr.K. Senthil Kumar[2*]

[1]Research Scholar, School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, India. shebapari.n2017@vitstudent.ac.in, https://orcid.org/0000-0002-8072-1347

[2*]Professor, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India. ksenthilkumar@vit.ac.in, https://orcid.org/0000-0001-6997-8398

Received: October 04, 2025; Revised: November 28, 2025; Accepted: January 02, 2026; Published: March 31, 2026

Abstract

Today, malicious actors often use social media to send, receive, and post false, misleading, or offensive content about other people. The effects of social media bullying on its targets are similar to those of threats, gossip, and harassment in the physical world. An alarming rise in mental health issues has resulted from cyberbullying, particularly in the younger population. The effects can include suicidal thoughts and low self-esteem. Several conventional machine learning techniques have been employed to automatically identify cyberbullying on social media. Although some traditional methods of machine learning have been designed to automatically identify cyberbullying, most of them are characterized by weaknesses in feature selection and classification rates. In order to deal with such issues, a new Termite Apriori Prediction Framework (TAPF) has been proposed in this paper. The model consists of a data collection and preprocessing phase to remove noise and a hybrid feature selection phase that combines Apriori rule mining with termite optimisation to select the most discriminative features. Cyberbullying is then classified with the help of these optimized features. It has been experimentally shown that TAPF performs better than the classical models such as the Logistic Regression, Naive Bayes, Long Short-Term Memory (LSTM), and the Support Vector Machine (SVM). The proposed structure has a recall rate of 96.2, precision of 96.7, F-score of 96.3, and an overall accuracy of 96.4, which is over 2 per cent higher than the best-performing baseline (SVM). Applied to a Python environment, TAPF offers an effective and dependable mechanism of detecting cyberbullying on social media sites.

Keywords: Cyber Bullying, Attack Prediction, Social Network, Cyber-attack, Pre-processing.

1 Introduction

Cyberspace is rapidly emerging as the predominant medium for comprehensive information, reflecting both its allure and difficulties. Serving as a pivotal gateway, cyberspace offers a wealth of information and resources from across the globe. People actively participate in vast social networks, utilising them extensively to interact with others, regardless of geographic and time differences. Each year, the use of social networks grows. However, information shared on social media is accessible to anyone, including individuals with malicious intent. As a result, many issues arise when it comes to the content shared on

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), volume: 17, number: 1 (March - 2026), pp. 392-410. DOI: *10.58346/JOWUA.2026.I1.022*

*Corresponding author: Professor, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India.

these platforms. Some users have the malevolent desire to harm others, leading to the emergence of cyber threats, which are incredibly difficult to combat. This study focuses on identifying deceitful accounts that intentionally cause harm or hide their true motives, as they pose a significant online risk to the general population. Corporate entities mitigate cyber risk through a combination of insurance coverage and preventative measures. Often, cyber threats are regularly monitored using resources such as global information platforms like Cvedetails.com. Cybersecurity encompasses various elements, such as strategies, tools, and protocols designed to safeguard cyberspace from threats and cyberattacks. In today's digital age, cybercrimes are proliferating at a rate that outpaces the capabilities of existing cybersecurity infrastructure in the realm of computers and information technology. A computer system is susceptible to risks due to inadequate system configuration, unskilled staff, and a lack of security measures. Thus, there is a pressing need for advances in cybersecurity techniques to combat growing cyber threats.

Attack methods are constantly advancing to breach systems and bypass traditional signature-based defenses. Similarly, security, web, and mobile technologies evolve to keep pace. Machine learning techniques, with their agility in adapting to novel scenarios, hold promise for addressing intricate challenges in the future, and they have already effectively addressed a range of issues in computer cybersecurity. Hence, machine learning methods are crucial in combating cybersecurity threats. Intrusion detection systems, virus detection, phishing detection, spam tracking, and fraud detection are some applications addressed by machine learning. Social networks have distinct security characteristics. Initially, they help users find data and communicate, which depends on trust. Typical users disclose large amounts of personal data, which may or may not be publicly accessible to their acquaintances. When the information is not public, its accessibility is governed by the trust network. In this scenario, a user grants permission exclusively to friends, allowing them to access the user’s personal information. Regrettably, social media platforms lack robust authentication methods, making it simple for impostors to infiltrate a user's network of acquaintances (Jbair et al., 2022). Moreover, in pursuit of popularity, users frequently accept friend requests from strangers, thereby exposing their personal details to unfamiliar individuals (Senthil Raja & Arun Raj, 2022). In the alternative scenario, the data in a user's profile is intentionally made public, as is the case on platforms like MySpace. Consequently, it can be accessed by anyone, regardless of their relationship with the user (Maftei et al., 2022). Despite extensive research on cybersecurity and machine learning-based detection methods, current approaches fail to detect attacks related to social networks. There is insufficient research available to detect attacks such as cyberbullying in social networks, especially hurtful online posts.

However, trust networks play a crucial role in ensuring security; they visually represent the mechanisms that safeguard users from unwanted contact by unauthorized entities (Jethava & Rao, 2024). Various levels of user awareness exist when it comes to potential threats, with the majority knowing about common threats that can arise (Bicchieri & Dimant, 2022). Twitter is a prominent social media network with over 500 million users (Hasan et al., 2024). Users share information through 140-character tweets, with around 500 million tweets posted every day (O’Regan, 2024). Twitter offers unique features, such as followership, usernames, profile images, locations, biographies, hashtags, and retweets (Aramburu et al., 2024). Approximately 80% of Twitter users tweet from their mobile phones (Howard-Sukhil, 2023). This media platform, however, has become a playground for cyberbullying (Zou et al., 2024). OSM not only connects individuals but also collects extensive data on human behaviour, whether textual or multimodal; these data serve as fuel for predictive models (Watson et al., 2024). Machine learning has proven to be effective in cyber threat detection, but current research has not focused on how feature extraction, noise handling, and hybrid frameworks can be used for predictions in large unstructured data, such as social network data. The current mechanisms of

moderation and manual reporting systems should not be expected to handle the huge amount, speed, and type of user-generated content on social media sites. Consequently, automated cyberbullying detection has been a very important area of research. The machine learning and deep learning methods that have been discussed include: Logistic Regression, Naive Bayes, Support Vector Machines and Long Short-Term Memory networks, which are useful in classifying online content into cyberbullying or non-cyberbullying. This is even though these approaches have significant limitations. Ineffective feature selection is one of the major problems. The text of social media is extremely unstructured, noisy and contextual, and may include slang, abbreviations, emojis and implicit semantic information. The traditional models either use manually engineered features or a high-dimensional feature space that may present redundancy and irrelevant features. This causes more false positives and worse generalization performance, especially in cases where one is to tell the difference between aggressive language and contextual or sarcastic utterances.

Additionally, most of the methods in place do not capture the associative trends of features that predict the behaviour of bullies in a satisfactory way. The deep learning models are powerful, but they frequently need big labelled datasets and the complementary high computational expenses, though the traditional classifiers might be missing important subtle semantic connections that are vital in the correct recognition.

The necessity of such limitations is encouraged by the desire to have a strong and efficient framework capable of (i) efficiently sampling discriminative features, (ii) pruning noise and redundancy, and (iii) enhancing the accuracy of classification without the burden of going overboard in computational costs. To overcome these issues, this paper suggests the Termite Apriori Prediction Framework (TAPF), which combines Apriori rule mining with termite Optimisation to improve the selection of features and prediction of cyberbullying. Through the realisation of associative feature relationships and optimized search procedures, TAPF is supposed to address the weaknesses of the existing models and deliver reliable cyberbullying detection that can be applied in real-life social media settings.

This study aims to bridge the mentioned gaps by proposing an enhanced framework using hybrid machine learning models to detect cyberbullying threats with greater accuracy.

The key contributions of this research work are as follows:

- The algorithm is initially trained using a cyberbullying database.
- Data are analyzed for noise, which is eliminated in the preprocessing phase, and then feature selection is performed.
- In the feature selection phase, Apriori rule mining is used to identify patterns, followed by termite optimization for feature selection.
- Based on the selected features, the cyber threat activities in the social network are predicted by a classifier model.
- The termite Apriori framework's performance is evaluated using performance metrics such as precision, accuracy, F-score, recall, and rate of error rate.

Online social networks have become increasingly popular in recent years. These platforms are attractive because of their varying options for online social networking and communication. Regrettably, they present privacy and security issues. Mughaid et al., (2023) created a new authentication-based machine learning method to overcome this issue. After the user enters a password, the method suggests using electronic face-processing verification as a second-factor authentication step. However, this method is not scalable across various social network platforms. As a result, social network platforms

have addressed the issue of bogus accounts, which continues to be one of the most serious challenges in social media on the internet.

Computer and information security often focuses on technology, viewing the human element in socio-technical systems as the most vulnerable element. However, these strategies fail to account for the cognitive attributes, requirements, and incentives of end users. Pollini et al., (2022) introduced a human factor maturity assessment in an innovative, holistic/Human Factors (HFs) approach to address this issue. This approach is utilised to evaluate the maturity level of pilot organisations regarding their capacity to confront and mitigate cyber threats and attacks. Theoretically, this technique is highly effective, but the practical prediction rate is much lower.

Facing an increasing number of cyberattacks, software companies are required to regularly implement new security measures. There is a growing need for heightened security measures in the industry. To overcome this problem, Jacob et al., (2022) have developed a Diffusion Convolutional Recurrent Neural Network (DCRNN) that can analyse spatio-temporal relationships, but when the encountered data have long sequences, the performance declines.

Cyberattack prevention and recovery face significant obstacles. For example, co-exploitation behaviour presents a problem because it uses multiple exploits simultaneously to take advantage of several software vulnerabilities. Yin et al., (2022) developed the Modality-Aware Graph Convolutional Network (MAGCN) technique to overcome these challenges. The technique utilises a reduced-dimensional feature space and merges map connectivity qualities with multimodal entity attributes and topologies to enhance the accuracy of connectivity forecasts. However, this has the drawback of being computationally costly. Table 1 summarises the research gaps addressed by the TAPF.

A review of previous studies reveals that some works give importance to network topologies, some to authentication, and some to detection. Very few studies address all these factors together (Agushaka et al., 2025). The common limitations of the previous studies are scalability, computational cost, noise sensitivity, and lack of a unified framework. These studies focus on classification and detection but do not extract associative patterns of malicious behaviours. Moreover, no framework uses Apriori-based association rule mining to uncover threat patterns, nor do they use multi-stage prediction using termite optimisation for cyberbullying threat detection.

The paper is organised as follows: Section II reviews the methods used in existing models and highlights research gaps. Section III presents the proposed model, Termite Apriori Prediction Framework (TAPF). Section IV discusses the results obtained and the ablation study. Finally, Section V concludes and summarises the paper along with the future scope.

2 Existing Models

Cybersecurity is critical to the continued progress of information technology and services. The rise in cyberattacks can be attributed to the accessibility of information shared through this technology. Today, cyber threats encompass a wide array of malicious software, including Trojans, viruses, botnets, and toolkits.

Using social media without falling victim to cybercrime is a challenging endeavour. Uploading personal content, such as photos, videos, and audio, can pose various security and privacy risks from potential malicious actors who misuse shared information. Conventional methods of preprocessing and feature extraction, as shown in figure 1, struggle to detect cyber threats when data have complex patterns or contain slang and sarcasm. This figure presents the limitations of traditional machine learning

approaches for detecting attacks or bullying, where preprocessing and feature extraction are performed manually. They use basic classification methods and miss important patterns, which results in low accuracy and unreliable predictions.

Conventional approaches are imprecise with poor value prediction accuracy rates. Consequently, the standard dataset estimation method is highly complex and potentially dangerous. In this study, a prediction algorithm framework is introduced to resolve this issue.

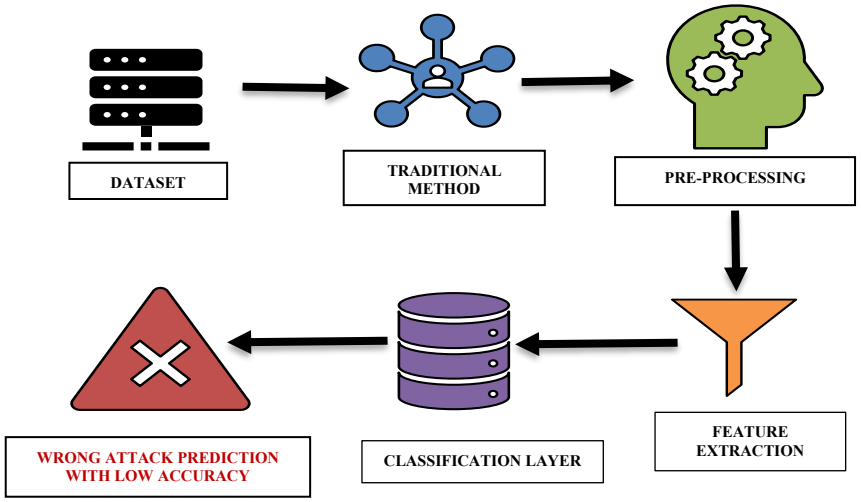

Figure 1: Conventional method of cyber threat detection and problem statement

3 Proposed Methodology

A new Termite Apriori Prediction Framework (TAPF) is implemented in this study. This proposed framework follows a multi-stage pipeline, as illustrated in figure 2, and uses a cyberbullying dataset collected from social media (Platform X, previously known as Twitter). The data are initialised and then passed to a preprocessing stage, where stop words and noise are removed, and numeric data are tokenised and normalized. In feature extraction, Apriori data mining is used to find frequent behavioural patterns associated with cyberbullying. The rationale for using Apriori data mining for the cyberbullying dataset is as follows:

- Cyberbullying data usually contains repeated linguistic patterns. Apriori data mining helps identify frequent itemsets of co-occurring words, phrases, and patterns in text.
- Apriori data mining extracts multi-word patterns. This helps capture the semantic structure of bullying language.
- Apriori data mining filters out rare and inappropriate word combinations and uses only substantial bullying patterns. After this stage, termite optimisation is implemented to refine Apriori association rules, removing noisy and irrelevant rules. Termite optimisation is chosen for the following reasons:
- It selects and refines the most predictive rules.
- It improves the feature quality and clarity using swarm-based reinforcement.

The optimisation function reinforces the classifier's unbalanced power by maximising useful patterns and removing noisy or redundant features, leading to a more accurate and context-aware cyberbullying detection model.

Finally, using the features extracted, the system generates a binary prediction of 1 for cyberbullying and 0 for non-cyberbullying activity. CatBoost is used as a classification model to handle the vast nature of social media data, which is highly unstructured (Agushaka et al., 2025).

Although more sophisticated NLP models like sentiment scores, TF-IDF embeddings, and transformer-based embeddings (e.g., BERT) yield richer linguistic semantics, the proposed framework has been designed with Apriori data mining to detect regular abusive behaviours, while termite optimisation reduces these behaviours into highly discriminative attributes. CatBoost also improves prediction using gradient boosting.

While the proposed design lacks deep linguistic embeddings, the optimisation-based feature engineering increases its discriminative power, and it identifies explicit cyberbullying patterns with high accuracy.

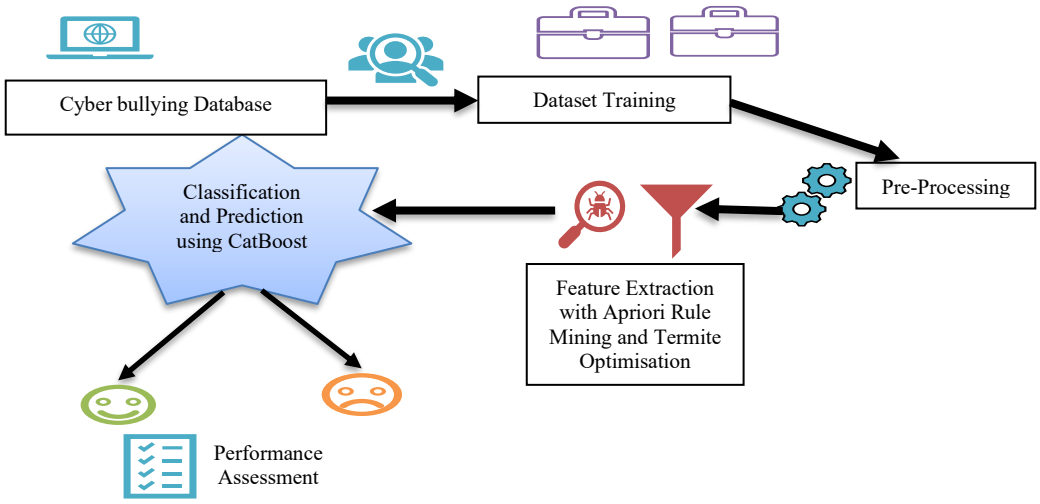

Figure 2: Proposed framework

Proposed TAPF Design

The designed TAPF incorporates termite optimisation (Minh et al., 2023) and the Apriori algorithm (Dasgupta & Saha, 2024). The planned framework has five different phases. The first layer is the training phase, where the data are imported. Noise is eliminated in the second-phase hidden layer. In the third phase, meaningful features are extracted, and classification is performed. The fourth layer applies the ideal fitness function. Lastly, the fifth layer provides the outcome of the cyberattack prediction.

- **Data Initializations and Training**

The data are initialised in the first step of the proposed TAPF process. Here, the framework imports the cyberbullying dataset. The framework is trained using an input dataset containing more than 47000 labelled cyberbullying tweets obtained from Kaggle https://www.kaggle.com/datasets/andrewmvd/cyberbullying-classification). The initialisation of the dataset is expressed in eqn. (1):

$$C(a_e) = a_e 1,2,3,4,5, \dots \dots ,, f \quad (1)$$

Where a_e is the initiated cyberbullying dataset, and $C(a_e)$ is the trained dataset factor and the number of datasets input. After the data are used for training, function preprocessing is initiated in the hidden layer to eliminate noise characteristics.

- **Data Preprocessing**

Preprocessing is a key stage of the prediction mechanism that improves the proposed TAPF's efficiency. The acquired text contains various noise, which distorts the data and further complicates the prediction process. Thus, before feature extraction, the noise is filtered out. A vital stage in data analysis is preprocessing, preparing raw data for subsequent analysis or modelling by organising, converting, and cleaning it. Eqn. (2) presents the noise removal function.

$$P(a_e) = |a_e(u, v) - a_e(w)| \quad (2)$$

Here, P is a preprocessing factor, u is a normal feature, and v represents noise particles; the noise features removed from the trained data are denoted as $|a_e(u, v)|$. Taking these steps can ensure that the data are reliable, comprehensive, and appropriate for constructing the predictive framework.

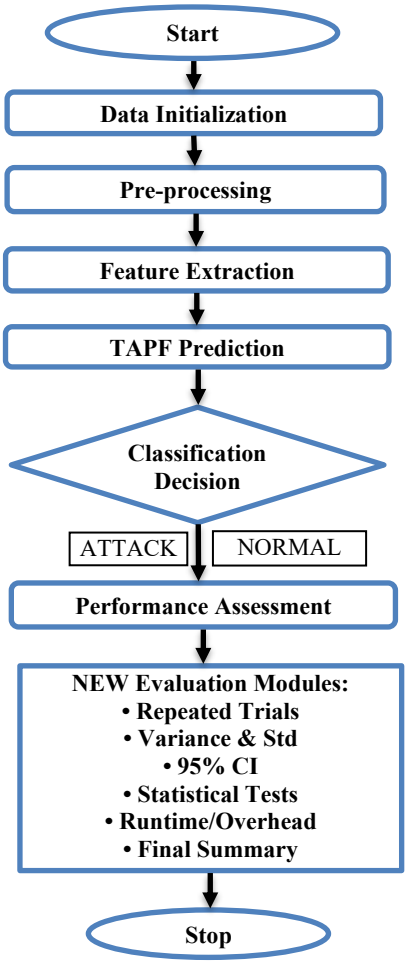

Figure 3: Workflow of the TAPF

- **Feature Extraction**

Feature extraction, an essential step in attack detection, is the technical name for dimensionality reduction. After eliminating redundant data, this phase outputs a feature representation set. The Apriori data-mining strategy is employed to determine which components in a dataset appear most frequently and which relationships are significant. Using this data-mining strategy, the proposed framework extracts the necessary attributes from the data. Cyberattack prediction relies heavily on feature extraction. In the cybersecurity context, feature extraction is the process of identifying and selecting relevant features or characteristics from data to aid in differentiating potential cyber threats from typical network behaviour (Gutiérrez-Batista et al., 2024). Eqn. (3) represents the feature extraction process.

$$H^*(a_e) = y + (r - t) \quad (3)$$

$H*$ refers to the Apriori feature prediction factor, y denotes an unwanted feature factor. Once unwanted features are eliminated, the extracted features move to the classification phase, where the attack is predicted.

- **Attack Prediction**

The features chosen in the feature extraction step are sent to the forecasting layer, where attacks are identified using a fitness method based on the termite attack mechanism framework. The attack-tracing function is defined in eqn. (4).

$$A(a_e) = l\left(\frac{I_{n(n,a)}}{P_r}\right) \quad (4)$$

Here, $A(a_e)$ denotes the termite attack-tracing score for candidate feature a_e, and $I_{n(n,a)}$ is the interaction or influence measure between node n and attribute a derived from termite-based feature evaluation. P_r is the pheromone-reinforcement factor, which represents how strongly the feature is supported by termite optimisation. $l(\cdot)$ is the linear transformation function used to scale the attack-score output. After the attack prediction, we move to the classification layer.

- **Classification Layer**

After prediction, the classification function is implemented. The CatBoost model is used to classify the features according to eqn. (5). The data are classified as normal activity or an attack. Here, the non-cyberbullying condition and cyberbullying conditions are 0 and 1, respectively, which are the termite fitness function range parameters.

$$K^*(a_e) = \{if\ (V_i = 0), non - cyberbullying\ \{if\ (V_i = 1), cyberbullying \quad (5)$$

The complete workflow of this framework figure 3 consists of the following steps:

- **Data initialisation:** The cyberbullying dataset is loaded and split into training and testing sets.
- **Preprocessing:** The text is cleaned and normalised, and noise, stop words, and unwanted particles are removed.
- **Feature extraction:** Apriori rule mining is applied for feature generation to extract discriminatory attack-related patterns.
- **Attack prediction:** A termite-based attack prediction framework generates prediction labels.

- **Classification phase:** The CatBoost algorithm classifies the data as an attack (0) or normal (1) behaviour.
- **Performance assessment:** Performance is assessed using accuracy, precision, recall, and F-score as metrics.
- **Repeated trials:** The model is executed for k = 10 independent runs.
- **Variation analysis:** After 10 runs, the metrics Mean, Variance, Standard Deviation, and 95% Confidence Interval (CIs) are calculated.
- **Statistical significance tests:** ANOVA testing is conducted to verify whether the TAPF performs significantly better than three or more existing models.

Algorithm: 1 TAPF

Input: Cyberbullying database D={d1,d2,...,dn}

Output: Classification layer C∈{Cyberbullying,Non-Cyberbullying}

Data initialization()

int a_e //individual data instance

//Dataset initialization

Pre-processing()

$P(a_e) \rightarrow |a_e(u,v) - a_e(w)|$

Filtering unwanted particles → v; $u \rightarrow normal\ particles$

//removing the noise factors

Feature extraction()

$H^*(a_e) = y + (r - t)$

//initialisation of feature extraction parameter

$featureextraction \rightarrow H * (a_e)$

// meaningful features were extracted by apriori data-mining function.

Attack Prediction()

$featureextraction \rightarrow H * (a_e)$

//initialisation of attack predicting factor

$\boldsymbol{A(a_e) \rightarrow l(l_n(n,a)}$

//attack is predicted by the termite attacking strategy function

Classification layer()

$if\ V_i = 0 \rightarrow non - cyberbullying$

attack

else(attack)

end for

$if\ V_i = 1 \rightarrow cyberbullying$

end for

$V_i \rightarrow estimating\ factor$

26 // Thus, attack is predicted by termite fitness function range.

27 end for

28 end while

29 Return $\boldsymbol{a_e}$

Algorithm 1 outlined the actions and operations indicated in the generated framework. The Termite Apriori Prediction Framework (TAPF) is based on associative rule mining combined with bio-inspired optimisation to identify cyberbullying. This algorithm starts with the initialisation and pre-processing of the dataset, in which normalisation, tokenisation, and removal of stopwords remove the noise and irrelevant data. Apriori rule mining is then used to obtain frequent and significant patterns of features according to the support and confidence pre-defined values. These characteristics represent associative links between terms that appear frequently in cyberbullying materials. Termite optimisation is also used to further improve the feature selection. At this phase, subsets of candidate features are considered as termite agents and fitted with a fitness function. Using iterative exploration and exploitation, the algorithm finds the best feature set, which will maximise the classification. At last, the optimised features are presented at the classification layer to classify content as either cyberbullying or non-cyberbullying. The Apriori rule mining and termite optimisation are combined to achieve better accuracy, less redundancy and better detection strength.

4 Results and Discussion

The cyberbullying data from the Kaggle website is used as the primary data source for implementing the TAPF. The acquired data are first preprocessed to improve their quality and eliminate any distracting elements. Additionally, the termite fitness approach is used to extract the effective characteristics and detect important features. Table 1 contains a list of prerequisites for carrying out the TAPF. After preprocessing, feature extraction is implemented using a hybrid approach. Apriori data mining identifies patterns that frequently co-occur with cyberbullying. Termite optimisation refines the output of step 1 by selecting the most useful features and removing redundancy.

Table 1: Description of the parameters

Parameter specifications	
Programming language	**Python**
Version	3.7.14
Datasets from	Kaggle
Dataset	Cyberbullying text data
Objective	Cyberbullying attack prediction
Optimization	Termite
Feature mining	Apriori
Python ID	Pycharm
Classifier	CatBoost

Case Study

The dataset employed in this research is derived from an extensive compilation of cyberbullying data, which was issued on April 15th, 2020 (https://www.kaggle.com/datasets/andrewmvd/cyberbullying-classification). The dataset is predominantly compiled from social media platforms, such as Kaggle, and contains an extensive range of cyberbullying incidents. In this study, the scope was restricted to Twitter data extraction (X), which yielded a dataset comprising more than 47,692 tweets explicitly categorised

as cyberbullying. The dataset is characterised by diverse demographic and compositional parameters, encompassing demographic variables including age, ethnicity, gender, and religion, as well as explicit labels denoting various forms of cyberbullying. This multiclass dataset is used to gain a comprehensive understanding of cyberbullying phenomena by identifying the interplay between diverse demographic characteristics and various forms of cyberbullying. The cyberbullying and non-cyberbullying data were classified according to age, ethnicity, gender, and religion. Table 2 describes the data used in the model.

Table 2: Data description

Total samples = 47692	
Cyberbullying	39747
Non-cyberbullying	7945
Training samples = 38153	
Cyberbullying	31832
Non-cyberbullying	6321
Testing samples = 9539	
Cyberbullying	7915
Non-cyberbullying	1624

The confusion matrix in figure 4 displays the proportion of accurate and inaccurate predictions for every potential relationship between the categorized data and the ground truth, or established reference data. The model correctly identified 1366 samples as Not Cyberbullying, which are true negatives (TN). The model incorrectly identified 258 samples as Not Cyberbullying, which are false positives (FP). Of all the Not Cyberbullying posts, 258 were incorrectly identified as Cyberbullying. The model missed 299 Cyberbullying posts, labelling them as Not Cyberbullying, which are false negatives (FN). The model correctly identified 7616 samples as Cyberbullying, which are true positives (TP). Here, the cyberbullying and non-cyberbullying predictions are represented by values of 1 and 0, respectively. The cell in the column and row that indicates the subclass was therefore given the percentages of samples from the data.

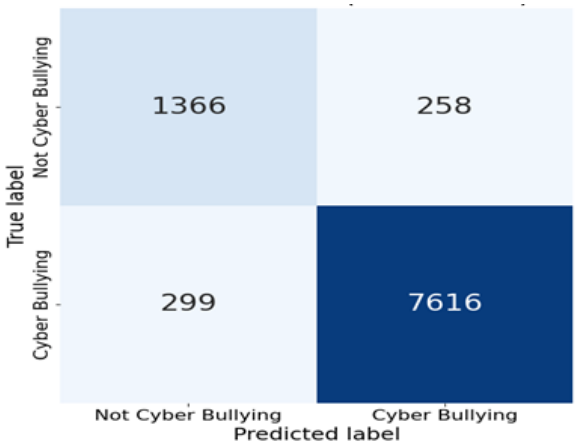

Figure 4: Confusion metrics for 9539 testing samples

Parameter Initialization and Experimental Settings

All the experiments were written in a Python environment to achieve model-to-model reproducibility and consistency. The dataset was also preprocessed before being classified using common preprocessing methods such as text normalization, removal of stop-words and tokenisation. Term-frequency-based representations were used to generate the feature vectors.

In the case of baseline classifiers, the parameterisation was done as follows. Logistic Regression was set to use L2 regularisation penalty and the maximum number of iterations of 1000 to guarantee convergence. The Naive Bayes classifier used a multinomial distribution and default smoothing parameters. The Support Vector Machine (SVM) made use of a radial basis function (RBF) kernel with a regularisation parameter.

The Long Short-Term Memory (LSTM) model was set with randomly fixed weights in the form of Xavier initialisation, the learning rate of 0.001, the batch size of 32, and the number of epochs set to 50 to prevent overfitting and early stopping.

In the proposed TAPF framework, Apriori rule mining was launched with a minimum support level of 0.3 and a minimum level of confidence of 0.6, which allowed the identification of strict associative rules among features. The termite optimisation algorithm started with a population of 30 termites, a limit of 100 executions and pheromone evaporation and deposition rates set empirically to a tradeoff between exploration and exploitation. The TAPF generated optimized subset of features was then employed to train the final classifier.

To evaluate the method's predictive capability, it was compared with various baseline models: Logistic Regression (LR), Naïve Bayes (NB), Long Short-Term Memory (LSTM), and Support Vector Machine (SVM). These models were used for comparison because they are the commonly used approaches in cyberbullying.

- Logistic Regression classifies instances as a 0 or 1, where 1 indicates success and 0 indicates failure. This model's main limitation is that it does not support the linear function representation of the value.
- Naïve Bayes is a simple method for creating an automated classification framework by assigning class labels to arrays of characteristic values and issue occurrences. It is a fundamental probabilistic classifier framework, but it struggles with complex contextual relationships.
- LSTM, based on an RNN, learns long-term dependencies. The issue of disappearing gradients can be optimally solved with an LSTM model. The main limitation is that LSTM is computationally expensive and sensitive to sequence length.
- SVM is highly effective in high-dimensional spaces but requires proper parameter tuning (Sudhakar & Kaliyamurthie, 2024).

Accuracy

Accuracy is the proportion of correctly identified cyberbullying instances out of all the instances discovered. The accuracy value is calculated according to eqn. (6).

$$A = \frac{T_P + T_N}{T_P + T_N + T_P + T_N} \tag{6}$$

Where T_P indicates the number of bullying instances that the framework correctly predicted as bullying, and T_N denotes the number of non-bullying instances that the framework correctly identifies as non-abusive. Although accuracy is a useful metric, it does not reveal additional information about the performance of the classifier in cases where the dataset is unbalanced. Here, the TAPF's accuracy is verified and contrasted with those of existing frameworks, namely, LR, NB, LSTM, and SVM, which obtained accuracies of 92.1%, 78.4%, 69.2%, and 94.3%, respectively. The accuracy of the validated

TAPF design is 96.4 %, which is higher than that of the baseline models. This increased accuracy confirms that the TAPF implementation performed well in predicting cyber threats.

Precision

Precision is defined as the ratio of the number of accurately recognised instances of bullying to the total number of bullying cases identified. The precision formula is expressed in eqn. (7).

$$P = \frac{T_P}{T_{PR}} \tag{7}$$

Here, the proposed TAPF's precision is verified and contrasted with those of existing frameworks, namely, LR, NB, LSTM, and SVM, which achieved precision values of 92.4%, 76.1%, 67.8%, and 95.1%, respectively. The precision of the TAPF is 96.7 %, which is higher than that of the baseline models, meaning that the TAPF minimises false positives more effectively.

Recall

Recall refers to the total number of bullying instances recovered from the entire bullying dataset. Eqn. (8) provides the recall formula.

$$R = \frac{T_P}{T_{AP}} \tag{8}$$

Here, the proposed TAPF's recall is verified and contrasted with those of existing frameworks, namely, LR, NB, LSTM, and SVM, which obtain recall values of 91.9%, 73.4%, 63.7%, and 94.9%, respectively. The recall of the TAPF model is 96.2 %, which is higher than that of the baseline models. This increased recall confirms that the TAPF implementation performed well in predicting cyber threats and yielded more relevant results.

F-Score

The F-score is used to assess how a classifier or framework performs when there is a need to balance precision and recall, as well as when the dataset is unbalanced and contains a significant proportion of real negative values. False positives and false negatives typically play a significant role in learning frameworks. To reduce the influence of actual negative values, the F-score assigns greater weight to these values. The formula for the F-score is expressed in eqn. (9).

$$F = 2 * \frac{precision * Re\,call}{Pr\,ecision = Re\,call} \tag{9}$$

The recommended TAPF's F-score result is confirmed and compared with those of LR, NB, LSTM, and SVM, which obtain F-scores of 92.1%, 71.8%, 47.6%, and 94.9%, respectively. The F-score of the TAPF is 96.3 %, which is higher than that of the baseline models. Table 3 shows the comparative performance of various baseline models used for cyberbullying detection, evaluated using Recall, F-Score, Precision, and Accuracy. This increased F-score number confirms that the TAPF implementation performed well in predicting cyber threats. Figure 5 compares all four metrics, accuracy, precision, recall, and F-score, between the TAPF and the existing frameworks, LR, NB, LSTM, and SVM.

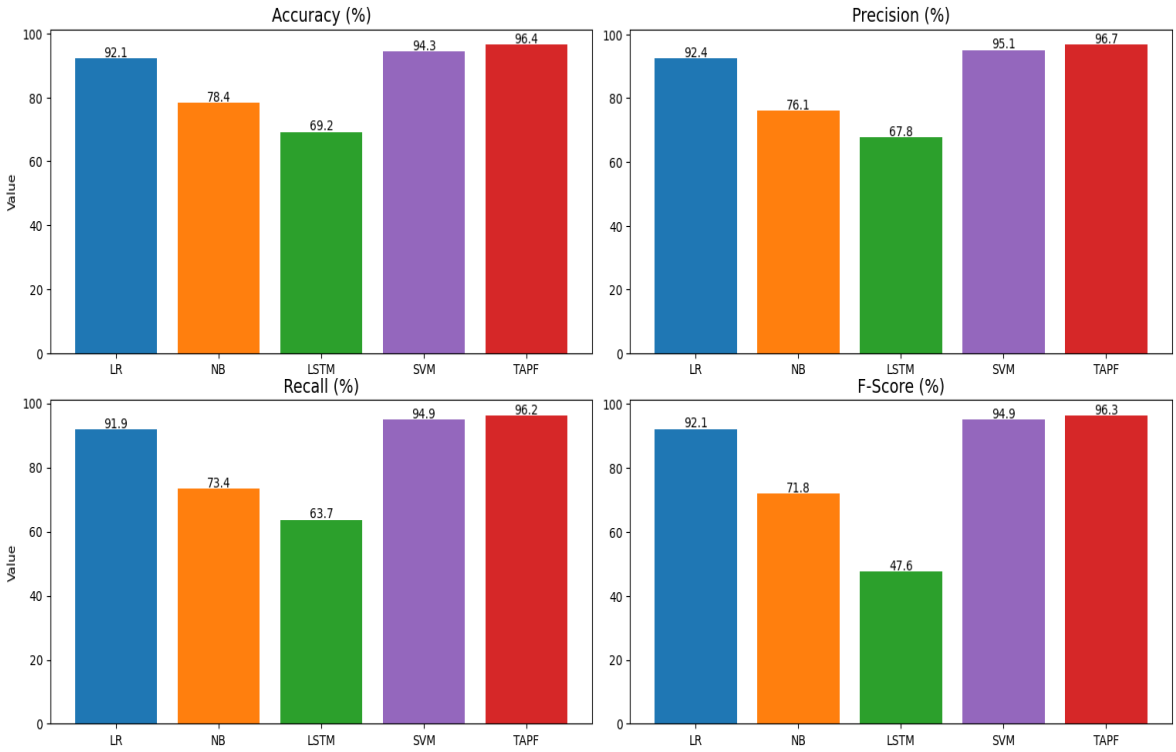

Figure 5: Comparative analysis

Error Rate

The error value is calculated to demonstrate the rate of efficiency of the model. Accuracy is subtracted from the total metric value to obtain the error value. Eqn. (10) expresses the error rate.

$$E_R = 1 - A \tag{10}$$

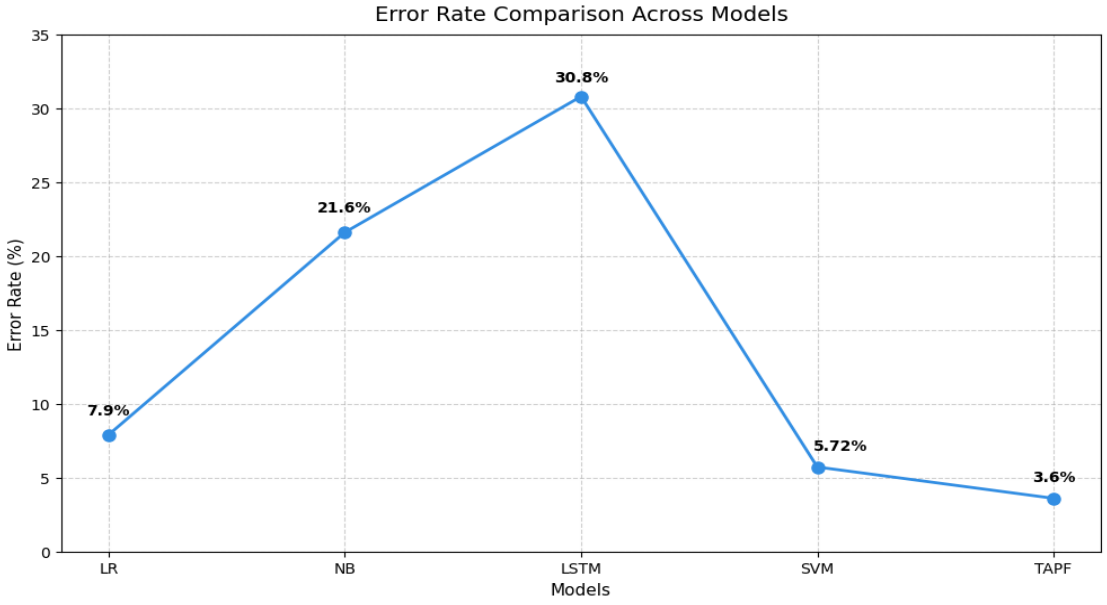

Figure 6: Comparison of error rates between methods

Figure 6 compares the error rate of the TAPF with those of LR, NB, LSTM, and SVM. Here, ER denotes the error rate factor. To assess the efficiency of the TAPF model, the error rate was computed alongside the other metrics. The proposed TAPF's error rate is compared with those of LR, NB, LSTM, and SVM, which achieved error rates of 7.9%, 21.6%, 30.8%, and 5.72%, respectively. The TAPF's error rate of 3.6% is lower than that of the baseline methods. Therefore, in the context of attack detection, the proposed framework has superior operational effectiveness.

Table 3: Overall performance statistics

Methods	Metrics			
	Recall	F-Score	Precision	Accuracy
Logistic Regression	91.9	92.1	92.4	92.1
Naïve Bayes	73.4	71.8	76.1	78.4
Long Short-Term Memory	63.7	47.6	67.8	69.2
Support Vector Machine	94.9	94.9	95.1	94.3
TAPF	96.2	96.3	96.7	96.4

5 Discussion

The results show that the TAPF is better at detecting cyberbullying than the baseline models. Apriori-derived rule sets are improved by termite optimisation, and CatBoost provides a robust classification. The model outperforms the baseline methods across all performance metrics.

Cross-Validation with Various Sets

In this study, phishing data were collected from authenticated open cybersecurity resources: the UCI Machine Learning Repository and the PhishTank (2024) repository, with 11,055 sample sites split into 5,485 legal URLs and 5,570 phishing URLs. After features were extracted and encoded, the dataset was about 2.3 MB, and it contained 30 numerical and categorical variables, such as URL length, the number of sub-domains, the validity of the SSL certificate, the age of the domain (how long ago it was registered), and the number of redirections. To measure the fairness of the performance of the model, the data were further divided into a 70:15:15 train-validation-test split. Training was performed using a fivefold cross-validation methodology, which enhanced the resilience of the model and minimised overfitting. The folds were trained individually and averaged to obtain the final measure, hence generalising the findings to unknown phishing patterns as well. All features were normalised to the scale [0,1]. Following the correlation analysis, instances of duplicated properties were eliminated, and then the data were fed to the TAPF model. Table 4 compares model performance on the phishing dataset.

Table 4: Comparison of performance on the phishing dataset

Fgjk'	Accuracy (%) ± CI	Precision (%) ± CI	Recall (%) ± CI	F-Score (%) ± CI	Error Rate ± CI	Execution Time (s) ± CI	p-Value (vs TAPF)
Logistic Regression	92.14 ± 0.38	91.80 ± 0.42	92.05 ± 0.40	91.92 ± 0.39	7.86 ± 0.38	0.42 ± 0.03	0.018
Naïve Bayes	89.72 ± 0.51	88.95 ± 0.56	89.30 ± 0.54	89.12 ± 0.55	10.28 ± 0.51	0.21 ± 0.02	0.007
LSTM	94.88 ± 0.34	94.30 ± 0.36	94.60 ± 0.35	94.45 ± 0.35	5.12 ± 0.34	6.84 ± 0.15	0.062
Support Vector Machine	95.42 ± 0.31	95.10 ± 0.33	95.25 ± 0.32	95.17 ± 0.31	4.58 ± 0.31	2.93 ± 0.08	0.041
TAPF	97.63 ± 0.27	97.40 ± 0.29	97.50 ± 0.28	97.45 ± 0.28	2.37 ± 0.27	1.12 ± 0.05	—

Ablation Study

A study of ablation was done to assess the personal performance of the main elements of the proposed Termite Apriori Prediction Framework (TAPF) and to compare the effects of these elements on the cyberbullying detection performance. Each of the modules is systematically isolated in the study to determine its effectiveness in enhancing the accuracy of classification and its strength. There were four experimental configurations that were taken into account: Baseline Classifier with no Feature Selection, in which all the extracted features were directly used as the input to classification. Apriori-Based Feature Selection Only, in which the Apriori rule mining was applied to identify frequent feature associations, and optimisation was not done.

Table 5: Ablation study of TAPF components

Configuration	Description	Recall (%)	Precision (%)	F-Score (%)	Accuracy (%)	Observations
Baseline Classifier (No Feature Selection)	All extracted features are directly used for classification	91.3	91.8	91.5	91.6	Performance affected by redundant and noisy features
Apriori Only	Feature selection using Apriori rule mining without termite optimisation	94.1	94.5	94.3	94.2	Improved associative feature identification, but limited feature reduction
Termite Optimisation Only	Feature selection using termite optimisation without Apriori rule mining	94.6	95.0	94.8	94.9	Better search-based optimization but lacks semantic association modelling
Full TAPF (Apriori + Termite Optimization)	Hybrid feature selection integrating Apriori and termite optimisation	96.2	96.7	96.3	96.4	Highest performance due to synergistic integration of associative mining and optimisation

Termite Optimisation Only, in which termite optimisation was used to select features with no associative rule mining. Whole TAPF Framework Apriori rule mining, and termite optimisation. The baseline classifier was found to have reduced performance because of the existence of duplicate and irrelevant features, indicating the weakness of direct usage of features on social media data with noise. Apriori rule mining alone showed performance improvement because it identified meaningful and frequent associations of features, but did not have an effective mechanism to reduce the set of selected features. On the same note, termite optimisation by itself improved the feature selection efficiency by searching the feature space, but was unable to contain semantic dependence among features. The full TAPF framework demonstrated the best performance in all assessment measures, thus proving that associative rule mining in combination with bio-inspired optimisation is a necessary tool in order to maximise detection rate. The obtained results of the ablation confirm that Apriori rule mining and termite optimisation cannot be used to realise the optimal performance on their own. Rather, their effect is synergistic and allows TAPF to address the issue of feature redundancy, information maintenance, and enhanced generalisation during cyberbullying recognition. Table 5 represents the ablation study of the proposed TAPF components.

6 Conclusions

Cyberbullying is an ongoing concern in the era of digital technology. It poses risks to the welfare of individuals, who are often targeted because of diverse characteristics, including religion, age, ethnicity, and gender. A novel TAPF was developed in this study to predict cyber threats in social networks. The proposed approach's performance was compared with Support Vector Machine, Logistic Regression, Naïve Bayes, and Long Short-Term Memory, with the proposed framework achieving 96.4% accuracy, 96.7% precision, 96.2% recall, 96.3% F-score, and a 3.6% error rate. Although these achievements are encouraging, there are a number of research directions that can be explored in future studies. To begin with, the structure can be scaled to the use of context-aware and transformer-based language models that would more effectively capture the semantic and syntactic peculiarities of the multilingual and code-mixed social media text. Second, the future work can be oriented on detecting cyberbullying in real-time by improving the computational efficiency of TAPF for large-scale streaming data. Third, emotional and sentiment-sensitive elements should be further included in the detection of weak and implicit bullying behaviour. As well, it would be beneficial to test the framework on cross-platform and cross-domain data sets to determine the generalisation abilities. Lastly, the use of explainable artificial intelligence (XAI) methods may increase the transparency of models and facilitate responsible and responsible usage in moderation systems in the real world. On the whole, the suggested TAPF is a solid basis for automated cyberbullying detection and has quite a number of potential opportunities for further research and practical implementation.

Acknowledgment

The authors would like to thank the department SCORE and SCOPE of VIT, University Vellore, for providing support and cooperation during this research. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Author Contributions

N. Sheba Pari—Conceptualisation, data curation, formal analysis, investigation, methodology, project administration, software, visualisation, roles/writing, original draft.

Dr.K. Senthil Kumar—Conceptualisation, formal analysis, project administration, supervision, validation, visualisation, writing—review and editing.

Funding

This research received no external funding.

Data Availability Statement

Datasets supporting the reported results can be found at

https://www.kaggle.com/datasets/andrewmvd/cyberbullying-classification

https://archive.ics.uci.edu/dataset/327/phishing+websites,

which are publicly archived datasets analysed or generated during the study.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

[1] Agushaka, J. O., Ezugwu, A. E., & Abualigah, L. (2025). A multi-strategy Horned Lizard Optimisation Algorithm for complex optimisation and advanced feature selection problems. *Journal of Big Data, 12*(1), 107.

[2] Aramburu, M. J., Berlanga, R., & Lanza-Cruz, I. (2024). A Data Quality Multidimensional Model for Social Media Analysis: MJ Aramburu et al. *Business & information systems engineering*, *66*(6), 667-689. https://doi.org/10.1007/s12599-023-00840-9

[3] Bicchieri, C., & Dimant, E. (2022). Nudging with care: The risks and benefits of social information. *Public choice*, *191*(3), 443-464. https://doi.org/10.1007/s11127-019-00684-6

[4] Dasgupta, S., & Saha, B. (2024). Big data analysis on medical field for drug recommendation using apriori algorithm and deep learning. *Multimedia Tools and Applications*, *83*(35), 83029-83051. https://doi.org/10.1007/s11042-024-18832-6

[5] Gutiérrez-Batista, K., Gómez-Sánchez, J., & Fernandez-Basso, C. (2024). Improving automatic cyberbullying detection in social network environments by fine-tuning a pre-trained sentence transformer language model. *Social Network Analysis and Mining*, *14*(1), 136. https://doi.org/10.1007/s13278-024-01291-0

[6] Hasan, M. A. U., Bakar, A. A., & Yaakub, M. R. (2024). Measuring user influence in real-time on twitter using behavioural features. *Physica A: Statistical Mechanics and its Applications*, *639*, 129662. https://doi.org/10.1016/j.physa.2024.129662

[7] Howard-Sukhil, C. (2023). Twitter & world literature: The development of hashtag communities as a global writing practice. *New Techno Humanities*, *3*(2), 90-100. https://doi.org/10.1016/j.techum.2024.02.001

[8] Ileri, K. (2025). Comparative analysis of CatBoost, LightGBM, XGBoost, RF, and DT methods optimised with PSO to estimate the number of k-barriers for intrusion detection in wireless sensor networks. *International Journal of Machine Learning and Cybernetics*, *16*(9), 6937-6956. https://doi.org/10.1007/s13042-025-02654-5

[9] Jacob, S., Qiao, Y., Ye, Y., & Lee, B. (2022). Anomalous distributed traffic: Detecting cyber security attacks amongst microservices using graph convolutional networks. *Computers & Security*, *118*, 102728. https://doi.org/10.1016/j.cose.2022.102728

[10] Jbair, M., Ahmad, B., Maple, C., & Harrison, R. (2022). Threat modelling for industrial cyber physical systems in the era of smart manufacturing. *Computers in Industry*, *137*, 103611. https://doi.org/10.1016/j.compind.2022.103611

[11] Jethava, G., & Rao, U. P. (2024). Exploring security and trust mechanisms in online social networks: An extensive review. *Computers & Security*, *140*, 103790. https://doi.org/10.1016/j.cose.2024.103790

[12] Maftei, A., Holman, A. C., & Merlici, I. A. (2022). Using fake news as means of cyber-bullying: The link with compulsive internet use and online moral disengagement. *Computers in Human Behavior*, *127*, 107032. https://doi.org/10.1016/j.chb.2021.107032

[13] Minh, H. L., Sang-To, T., Theraulaz, G., Wahab, M. A., & Cuong-Le, T. (2023). Termite life cycle optimizer. *Expert Systems with Applications*, *213*, 119211. https://doi.org/10.1016/j.eswa.2022.119211

[14] Mughaid, A., Obeidat, I., AlZu'bi, S., Elsoud, E. A., Alnajjar, A., Alsoud, A. R., & Abualigah, L. (2023). A novel machine learning and face recognition technique for fake accounts detection system on cyber social networks. *Multimedia Tools and Applications*, *82*(17), 26353-26378. https://doi.org/10.1007/s11042-023-14347-8

[15] O'Regan, G. (2024). Ethical social media. In *Ethical and legal aspects of computing* (pp. 117–135). Springer. https://doi.org/10.1007/978-3-031-52664-0_6

[16] Pollini, A., Callari, T. C., Tedeschi, A., Ruscio, D., Save, L., Chiarugi, F., & Guerri, D. (2022). Leveraging human factors in cybersecurity: an integrated methodological approach. *Cognition, Technology & Work*, *24*(2), 371-390. https://doi.org/10.1007/s10111-021-00683-y

[17] Senthil Raja, M., & Arun Raj, L. (2022). Detection of malicious profiles and protecting users in online social networks. *Wireless Personal Communications*, *127*(1), 107-124. https://doi.org/10.1007/s11277-021-08095-x

[18] Sudhakar, M., & Kaliyamurthie, K. P. (2024). Detection of fake news from social media using support vector machine learning algorithms. *Measurement: Sensors*, *32*, 101028. https://doi.org/10.1016/j.measen.2024.101028

[19] Watson, E., Viana, T., & Zhang, S. (2024). Machine learning driven developments in behavioral annotation: A recent historical review. *International Journal of Social Robotics*, *16*(7), 1605-1618. https://doi.org/10.1007/s12369-024-01117-1

[20] Yin, J., Tang, M., Cao, J., You, M., Wang, H., & Alazab, M. (2022). Knowledge-driven cybersecurity intelligence: Software vulnerability coexploitation behavior discovery. *IEEE transactions on industrial informatics*, *19*(4), 5593-5601. https://doi.org/10.1109/TII.2022.3192027

[21] Zou, W., Drake, A. P., Masur, P. K., Whitlock, J., & Bazarova, N. N. (2024). Examining learners' engagement patterns and knowledge outcome in an experiential learning intervention for youth's social media literacy. *Computers & Education*, *216*, 105046. https://doi.org/10.1016/j.compedu.2024.105046

Authors Biography

N. Sheba Pari, she is a PhD Research Scholar in the School of Information Technology & Engineering (SITE) in VIT, Vellore, India. Her research interests include Information and Cyber Security, Social Networks, Artificial Intelligence and Network Engineering. She received her M. Tech in Software Engineering from Visvesvarya Technological University, Belgaum, India and B.E. in Information Science & Engineering from Visvesvarya Technological University, Belgaum, India. She has over 13 years of teaching experience as an Assistant Professor in various Engineering colleges.

Dr.K. Senthil Kumar received his Ph. D degree from the Department of Computer Science and Engineering at Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India, in the year 2016. He is currently a Professor in VIT Vellore India. His research interest areas include Knowledge in Data Engineering, Mobile Computing, Software Engineering, Mobile Data Management, Wireless Networks, Cloud Computing, Data Science, and Machine Learning.