National College of Ireland

Lightweight Deep Learning Approaches for Efficient Detection of Cyber-Attack Text

MSc Research Project
Data Analytics

Hemanth Ponduru
Student ID: 23416297

School of Computing
National College of Ireland

Supervisor: Cristina Hava Muntean

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Hemanth Ponduru ……. ……
Student ID:	23416297 ……
Programme:	MSc Data Analytics …… **Year:** 2025 ……..
Module:	MSc (Research) Practicum Part 2 ……
Supervisor:	Cristina Hava Muntean ……
Submission Due Date:	11/12/2025 ……
Project Title:	Lightweight Deep Learning Approaches for Efficient Detection of Cyber-Attack Text ……
Word Count:	8345 …… **Page Count** 21 …….

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	P.Hemanth ……
Date:	10/12/2025 ……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You should make sure that you have a HARD copy of the project as a reference to yourself and also in case the project gets lost or mislaid. It is not sufficient to have it in a computer.	□

The assignments will be dropped in the assignment box outside the office to the Programme Coordinator Office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Lightweight Deep Learning Approaches for Efficient Detection of Cyber-Attack Text

Hemanth Ponduru
23416297

Abstract

The massive proliferation of connected devices across the Internet of Things, Industrial Control Systems, and social media platforms has significantly increased vulnerability to advanced cyber-attacks, including malware, phishing, and botnet intrusions. Traditional rule-based systems and existing machine learning or deep learning models, despite their detection capabilities, often demand high computational resources and are unsuitable for real-time or resource-limited environments. These limitations underscore the need for models that maintain strong predictive performance while operating efficiently on constrained infrastructures. To address this challenge, this study investigates lightweight attention-based models for efficient and cyberattack text detection. The research leverages a compact LSTM-based neural architecture enhanced with a single attention mechanism that identifies the most relevant threat indicators while minimizing unnecessary computation. This selective attention enables effective operation within IoT and edge environments, where storage, processing power, and latency are critical constraints. The study implements a full pipeline encompassing data preprocessing, feature extraction, attention-driven weighting, and lightweight classification. Performance evaluation using accuracy, precision, recall, F1-score, and computational efficiency demonstrates that the lightweight LSTM + Attention model achieves superior results, reaching 94.31% accuracy and an F1-score of 94.33%, outperforming all traditional ML and standalone DL models. Overall, this study contributes of scalable, reliable, and resource-efficient cybersecurity solutions capable of adapting to the evolving complexity of modern cyber threats through the use of attention-enhanced lightweight architectures.

1 Introduction

The cyberattacks have made the digital infrastructures very vulnerable due to the massive growth in the number of connected devices in the Internet of Things (IoT), Industrial Control Systems and social networking sites. Such systems generate huge amounts of unstructured information that can harbour high-level threats such as malware, phishing and botnets attacks. The conventional cybersecurity tools which are founded on fixed signature and rule-based logic are no longer adequate to combat the dynamic attack vectors which evolve to overcome defines tools. As a result, the field of cybersecurity has shifted to the realm of data-driven solutions, with the machine learning (ML) and deep learning (DL) models having demonstrated a high potential to detect abnormalities in the large-scale network traffic and online interactions (Mishra et al., 2022). Nevertheless, with the rise in the sophistication of cyber threats and the sustained exponential growth of datasets, the necessity of efficient, 1 scalable, and resource-conscious models, in particular, in edge and IoT settings, has been of utmost importance.

However, even though impressive advances in ML and DL have occurred in the field of intrusion detection and threat classification, the majority of the current methods tend to be extremely resource-intensive in terms of computational power and large labelled datasets to reach the high level of accuracy. These issues render traditional models inappropriate to be deployed in real-time in settings with a low amount of memory, processing, or energy access. Moreover, most of the existing architectures compute all features equally, which causes redundancy and low efficiency in detecting cyber-threats. It is thus urgently needed that a new generation of models be developed that can be used to detect crucial patterns in cyber data at minimum resource consumption. Slim models that focus on attention have also become an exciting solution to fill this gap since they enable the model to selectively highlight the useful features of network or textual data, maximizing processing efficiency as well as detection (Alkhudaydi, Krichen, and Alghamdi, 2023; Nguyen and Kebande, 2024). The reason behind this research is that these models can offer effective cybersecurity with the cost of the computational burden of conventional deep architectures. In this paper, a lightweight Long Short-Term Memory (LSTM) model with the attention mechanism was employed as the main structure of cyberattack text classification.

The current researches have focused on developing powerful ML and DL models to identify threats and forecast cyber threats. One of them is the ensemble-based ML systems proposed by Okey et al. (2022) to detect cyberattacks on IoT networks, and Recent studies have considered the application of advanced deep learning architectures to bolster malware and attack detection in IoT settings. As an illustration, Alkhudaydi et al. (2023) used LSTM, GRU, and RNN to learn sequential patterns in the traffic of IoT networks. Their research proved that deep learning models are more effective in detecting complex behavioural marks of cyberattacks in comparison with the conventional machine learning techniques. To detect malware in IoT. On the same note, Al-Quayed, Ahmad, and Humayun (2024) offered a predictive framework that combines the use of ML and DL models to protect Industry 4.0 wireless networks, and Vitorino, Silva, Maia, and Praca (2025) discussed the idea of strong feature selection to improve the performance of attack detection. Although these models are very useful in enhancing the accuracy of classification, they have high computational cost, complexity of the model, and low scalability when applied in real time. Contrarily, lightweight attention-based models can also preserve the same level of defectiveness with reduced processing overhead, making them one of the possible solutions to the limitations found in previous research (Mubeen et al., 2025).

Research Question and Justification

Research Question:

How can lightweight model such as attention-based LSTM enhance the efficiency of cyberattack detection while maintaining high detection accuracy?

The rationale behind this research question is that there is an increasing need to have cybersecurity systems that are flexible in terms of performance and computational capabilities.

Since existing ML and DL architectures have resource constraints in resource-constrained settings, this paper will seek to understand how attention mechanisms can be applied through lightweight neural architectures to optimize threat detection. The research can help to promote effective and flexible cybersecurity measures in line with contemporary network infrastructures by answering this question.

This study aimed to examine how effective lightweight attention-based LSTM models are with respect to cyberattack text detection when the amount of computational resource is limited. The paper aimed at investigating how attention mechanisms have the potential to emphasize threat-relevant features more effectively than the traditional deep learning models. Other goals were to evaluate the performance of the lightweight framework based on the standard evaluation metrics and to establish whether it can be used in the IoT and edge environments.

2 Literature Review

Three primary research directions that are applicable in this study have been reviewed in the following subsections. To begin with, the current machine learning methods of cyberattack detection are analysed to find out how the conventional algorithms address the threat classification. Second, the methods based on deep learning are mentioned, and the focus is on their capacity to represent the intricate patterns in cybersecurity data. Third, hybrid and ensemble methods are discussed to put emphasis on the advantages of using more than one model. A combination of these domains gives the basis of determining the research gaps that lightweight LSTM-Attention architectures can solve.

2.1 Machine Learning Models in Cybersecurity

Application of machine learning models in contemporary cybersecurity systems has been a critical component of a cybersecurity system because it enables identification and categorization of threats using intelligent data analytics. In contrast to conventional intrusion detection systems that rely on fixed signature or rule-based reasoning, ML models are resilient to changing threat patterns by learning the patterns in network traffic data. This adaptive learning process enables ML to identify complex and subtle anomalies that are usually not detected using the traditional means. Mishra et al. (2022) highlighted that the Support Vector Machines, Random Forests, Logistic Regression, and Decision Trees are the possible ML methods that can be used to detect cyber threats in the Internet of Things (IoT) setting. On the same note, Al-Quayed, Ahmad, and Humayun (2024) proved that these ML algorithms as Decision Tree and Multi-Layer Perceptron can be implemented into predictive systems to detect intrusion in Industry 4.0 wireless sensor networks to supplement proactive threat detection and prevention. All these works suggest that ML-based solutions are not only flexible but can contribute to the overall enhancement of the quality and effectiveness of cybersecurity systems, in general.

Besides network intrusion detection, other types of cyber threat analysis, which have been effectively implemented using ML models, are phishing and anomaly detection. Benavides-

Astudillo et al. (2023) used Natural Language Processing (NLP) and ML to identify the phishing attack by using text data on the websites of the websites used by the fraudsters. The Okey et al. (2022) building an ensemble classifier called BoostedEnML is an integration of various algorithms of the ML to improve the power and credibility of identifying various types of cyberattacks on IoT systems. In addition, Nguyen and Kebande (2024) benchmarked various ML classifiers, including Random Forest, Logistic Regression, and Quadratic Discriminant Analysis on malware data, and have found out that the ensemble and advanced ML classifiers will always be more efficient than the traditional statistical ones. These results underscore the fact that ensemble learning and hybrid ML models are scalable, flexible and resistant to changing cybersecurity threats.

Other than detecting intrusion and phishing, digital abuse and other malicious activities on the internet like cyber bullying and hate speech have also been dealt with by ML applications. In the article Alqahtani and Ilyas (2024), the authors suggested a hybrid ML model, which was a combination of various standard algorithms and was used in identifying aggressive content on the internet, which demonstrates the promise of ML in the context of social media moderation. Similarly, Mubeen et al. (2025) used stacked ensemble classification model whereby a chain of ML models was used to find hate speech in social sites. Akar (n.d.) also made a comparison of the effectiveness of various ML classifiers in detecting cyberbullying grounded on Natural Language Processing and emphasized the significance of the algorithmic selection in the successful outcomes. All these studies demonstrate that ML methodologies can play a significant role in minimizing various types of cyber threats, beginning with network-level attack, social and psychological by applying data-driven intelligence and adaptive patterns recognition.

2.2 Deep Learning Model in Cybersecurity

Deep learning has already become a groundbreaking technology in cybersecurity, and its advanced capabilities of identifying, classifying, and predicting advanced cyber threats. Unlike the existing machine learning designs, which employ features that are handcrafted, the DL models are modelled to discover hierarchical features in large and heterogeneous data sets, thereby enabling softer detection of threats. This model demonstrated that the DL systems would be capable of learning complex time-dependent relationships within network traffic to identify anomalies that could be a sign of malicious activity. Similarly, Alkhudaydi, Krichen, and Alghamdi (2023) proposed a predictive framework of cybersecurity attacks in the IoT environment in terms of the DL, in which the models of LSTM, Gated Recurrent Unit (GRU), and Recurrent Neural Network (RNN) were utilized to determine the most significant features of the network traffic data. Their results supported the hypothesis that the DL models, especially their sequential versions, can learn the dynamics of cyberattacks more efficiently than the traditional ML models can.

The versatility of DL models can be transferred to cyber-physical systems (CPS) as well, and the identification of malicious intrusions is essential to ensure the reliability of the operations. Utubor (2023) discussed gradient boosting-based DL models, i.e., XGBoost and LightGBM,

to identify attacks in CPS networks and the ways in which these models can be improved with the help of proper sampling and feature selection strategies. Such DL-based systems do not only identify known threats, but also evolve to novel attack vectors which is important in dynamic industrial and IoT settings. Likewise, the importance of quality data and feature selection in the improvement of the adversarial robustness of the DL-based cybersecurity systems was investigated by Vitorino, Silva, Maia, and Praca (2025). They discovered that the DL architectures can be successfully generalized to adversarial settings when they are trained on trustful and mixed data, and this can reduce the vulnerability to evasion and perturbation attacks. These papers indicate that there is a necessity to combine advanced methods of DL with data optimization processes in order to possess robust and effective cyber defines mechanisms.

Besides network intrusion detection, DL has also been applied in resolving content-based security threats such as detection of phishing and cyber bullying. Textual characteristics of suspicious web pages were examined using deep learning and the Natural Language Processing (NLP) system with the introduction of such models as LSTM and GRU to detect phishing attacks in Benavides-Astudillo et al. (2023). On the same note, Aliyeva and Yağanoğlu (2025) used Multi-Layer Perceptron (MLP) architecture in detecting cyberbullying in Turkish Twitter text, and the application shows how DL can be used to process linguistic and contextual information. Another study by Purba et al. (2024) also related to this one, as it develops a modification of LSTM that is specifically tailored to detect the offensive use of language in the comments on the Indonesian social media, which indicates the intercultural and interlingual flexibility of DL. All these studies suggest that the DL architectures and their capability to learn deep representations and comprehend the context present a powerful weapon to defeat different cybersecurity risks, including IoT network intrusions and social media-based abuses.

2.3 Hybrid and Ensemble Approaches for Enhanced Cybersecurity Intelligence

Due to the constantly increasing complexity of cyberattacks, single-model frameworks, whether machine-learning models or deep-learning models, are unlikely to provide the comprehensive threat detection and resilience. To overcome these deficiencies, researchers have turned to hybrid and ensemble methods which combine the strengths of two or more algorithms or paradigms of learning. The techniques exploit the diversity of the classifiers to reduce false alarms, generalization and handle imbalanced or high-dimensional data. To provide an example, Okey et al. (2022) suggested the BoostedEnML framework, which is an ensemble-based intrusion detection model, which uses boosted learning algorithms, such as LightGBM and XGBoost in combination with other base classifiers. They have demonstrated that the combination of models can lead to the improvement of the capacity of detection in various attack scenarios of the IoT in a computationally efficient manner. On the same note, Nguyen and Kebande (2024) pointed out that the ensemble-based models such as the Random Forest are better than the individual models since they integrate the decision of different weak learners to form a stronger predictive model. All these findings confirm the fact that

hybridization provides greater adaptability to models and provides resilience in dynamic and heterogeneous network settings.

Hybrid systems go beyond the mere aggregation of models since they combine both ML and DL systems to deal with multi-faceted cybersecurity issues. Al-Quayed, Ahmad, and Humayun (2024) suggested a predictive intrusion detection system of Industry 4.0 networks that uses both ML algorithms to classify and the DL elements to extract intelligent features at the same time. This is a hybridized method that fills the gap between data-driven automation through the use of ML to explain and DL to abstract. Similarly. It was pointed out that hybrid deep architectures can be used to simulate different threat scenarios and synthesize data to train, which improves resistance to unknown attack patterns. The synergies between these systems are a pointer of the transition to the intelligent, context sensitive security systems that will be able to self-evolve in response to the emerging cyber threats via a continuous learning and feedback process.

Furthermore, the ensemble and hybrid approaches have also been found to be applicable in the anthropocentric domain of cybersecurity like social media monitoring and behavioural analysis. As illustrated by Mubeen et al. (2025), a stacked ensemble model is a set of several ML algorithms that identify hate speech and abusive content on the Internet and ensure that the process of identifying text-based threats becomes more precise and context-dependent. In a similar study, Aliyeva and Yağanoğlu (2025) used a hybrid deep learning model to detect cyberbullying on Twitter, which covers text and social media aspects, and this tendency can be explained by the fact that multimodal data sources are gaining popularity. On the same note, Akar (n.d.) also emphasized the role of feature engineering and algorithmic combination in detecting cyberbullying relying on Natural Language Processing (NLP). All these works point to the fact that ensemble and hybrid strategies are not limited to network-level security but can be applied to the protection of digital interactions and online communities. Combining various paradigms of learning, hybrid cybersecurity systems do not only offer greater detection potential but also offer adaptive intelligence in line with the changing sophistication of cyber threats in the digital era.

2.4 Research Gaps

Despite the widespread studies that have been performed with the application of machine learning, deep learning, and hybrid models to cybersecurity, multiple limitations are still present in current frameworks. Most of the studies reviewed are mainly concerned with the high performance of detection but do not pay much attention to the computational efficiency, scalability and applicability in real-time. As an example, where Alkhudaydi, Krichen, and Alghamdi (2023) suggested more sophisticated deep-architecture malware and IoT attack detection models, they use large datasets and demand substantial computing power, which restricts their application to IoT devices with limited resources. Ensemble methods like those in Okey et al. (2022) and Mubeen et al. (2025) are also better at detection robustness, but they are more likely to add complexity and inference time to a model and introduce latency in a real-world detection system. Also, the issue of of DL models is highly problematic, since black-

box models complicate the process of explanation to the analysts, why some network behaviours are considered threats. All these problems point to the necessity of efficient, and lightweight models that can be used to detect cyber-threats with a high degree of accuracy at a low cost of computation and provide a transparent decision-making.

In order to address these limitations, scientists are starting to consider the possibility of lightweight attention-based models, that is, attention mechanisms with the efficiency of a small neural architecture. Attention mechanisms allow the model to be selective to the most relevant features in the input data and thus enhance accuracy without the need to compute on a large scale. To implement the attention modules into lightweight frameworks or Transformers with reduced depths, cybersecurity models can be trained to concentrate on significant features of traffic or textual clues and ignore the redundant data. This selective focus allows the efficient acquisition in low-power or edge computing environment. Additionally, the factor that makes the attention-based architectures explainable is the fact that the attention weights are visually represented as the features that lead to a particular decision. This transparency can not only be used to build trust between cybersecurity practitioners, but also be used to audit models and comply with sensitive applications, including industrial internet of things and digital forensics.

The efficacy and flexibility can form the basis of real-time cyber defines since the attention mechanisms can be deployed into lightweight architectures. As an example, an attention-based model might be considered, which examines network traffic and dynamically assigns significance to packets or packet streams with unusual patterns, and minimizes training and false positives. Similarly, in text-based applications, such as phishing or cyberbullying detection, attention-based NLP models can attract attention to contextually relevant words, and systems can be more effective distinguishing between harmless and malicious content. Transfer learning can also be applied to these models to make them adapt quickly to changing threat landscapes without re-training them entirely. Therefore, lightweight attention-based models are adopted to fill the gap between the performance and practicality, providing a channel through which scalable, real-time, and explainable cybersecurity solutions can be developed to overcome the weaknesses of the current ML and DL strategies.

3 Research Methodology

3.1 Research Design Using CRISP-DM Methodology

As the methodology of the study, CRISP-DM (Cross-Industry Standard Process of Data Mining) was used to present a systematic and iterative process of implementing that will be aligned with the goals of cybersecurity. The Business Understanding stage formulates the issue of high computational cost. Data Understanding and Data Preparation stages include the collection, cleaning, tokenization and formatting of network and textual data to be used in lightweight architectures. The Modelling phase involves the incorporation of attention modules into small neural structures to improve selective feature learning and decrease the computational cost. The Evaluation stage aimed at measuring efficiency, explainability, and

adaptability instead of the conventional accuracy measures, and making the model fit the requirements of the real-time cyber defines. Lastly, the Deployment stage considers the model implementation into the IoT and edge computing systems that would allow constant monitoring and responsive adaptation to changing threats via retraining and feedback loops. This systematic CRISP-DM alignment assists of an effective, understandable, and scalable cybersecurity framework.

3.2 Dataset Description

The dataset used in this study contains labelled cybersecurity-related text instances that describe various malicious and benign activities, dataset sourced from link (Ramoliya et al,2023). Each record includes a short textual description along with a corresponding class label indicating whether it represents an attack or a non-attack event. This dataset provides a suitable foundation for training and evaluating machine learning and deep learning models for cyberattack text classification.

The data set in this study is the labelled textual data on cybersecurity incidents, in particular, various entities and patterns of malicious activities. Each record has several attributes like index, text, entities, relations, comments and respective labels which point to certain cyber-related terms like malware, attack-pattern, and time. The dataset format facilitates the recognition of entities by giving the start and end positions of each annotated label in the text, which allows mapping exact textual content and threat indicators. As an example, sentences that explain the behaviour of attacks or mechanisms of infection are annotated using appropriate entity tags that can help understand the context. These annotated data were essential in training and testing the lightweight attention-based models because they enabled the model to learn how to selectively focus on critical threat indicators in sentences. Through this organized data, the study will guarantee that the model that is able to understand semantic connections in the texts of cybersecurity and be computationally efficient throughout the process of extracting features and classifying them.

3.3 Data Preprocessing

Preprocessing of data is an important step in making sure that the data is clean and structured and can be used to implement machine learning-based cyberattack detection. The raw data had initially several columns which comprised identifiers, entity annotations, and relational data. Nonetheless, to construct a lightweight text-based detection model, two important columns were kept, namely, label and text. This choice was made to keep the information at hand that is relevant and the column with the text was the input data and the label column was the indication of whether the text was an attack or a benign case. The label column had all data points with non-attack classes that were substituted with the word benign to manage the inconsistencies and completeness of all the records. To create a classification setup, all non-benign entries were categorized as attack, providing a balanced and structure for downstream modelling.

Once the label column was refined, the additional data cleaning was done by deleting redundancy through deleting duplicates and other irrelevant features. The presence of duplicate entries that may cause overfitting and bias during model training was detected and eliminated entirely so that only unique and meaningful samples were left in the dataset. Duplicates were then discarded and each label was coded numerically where benign was coded as 0 and attack as 1. This conversion was necessary to be compatible with machine learning algorithms which need numerical values to classify data.

The second step was aimed at intensive text preprocessing to prepare the input data to extract features. Any textual material was turned to lower case to ensure consistency and any noise like URLs, email addresses, numbers and special characters were filtered out systematically to leave only significant linguistic elements. Stop words are words that are usually common and they do not add much semantic meaning to the words and hence they were removed to reduce the dimensions and bring out the key contextual words. Lemmatization was also used to bring the words up to the base forms so that the model could learn to treat the variations of a word as one. This step had the effect of reducing the vocabulary and enhancing the features extracted.

Lastly, the processed textual data were converted to numerical feature representations with the help of the Term Frequency -Inverse Document Frequency (TF-IDF) vectorization technique. TF-IDF allocated a weight to words and bi-grams depending on their significance and frequency within the dataset, which contained important linguistic information applicable in differentiating between text associated with an attack and non-attack text. In order to balance the evaluation of the models, the dataset was then split into training and testing subsets with the help of a stratified split, preserving the original ratio of attack and benign samples. Such preprocess activities were systematized in nature and transformed the data into clean structured and information rich format that may be used in formulating and analysing models.

3.4 EDA (Exploratory Data Analysis)

The exploratory data analysis provided the opinion of the composition and textual characteristics of the dataset. The label distribution plot showed that the dataset was healthy with regard to the amount of benign and attack cases, which gave equal conditions to learn. The histogram of text length revealed that the length of most of the samples was short in nature, which is the feature of a cyber threat notification and intelligence brief. The keywords which were dominant in the visualization of the word cloud were attack, malware, system, file, and user which are the key themes that are present throughout the data. A set of these visual indicators helped to make the conclusion that the dataset was contextually valid and could be used to train effective cyberattack detection models.

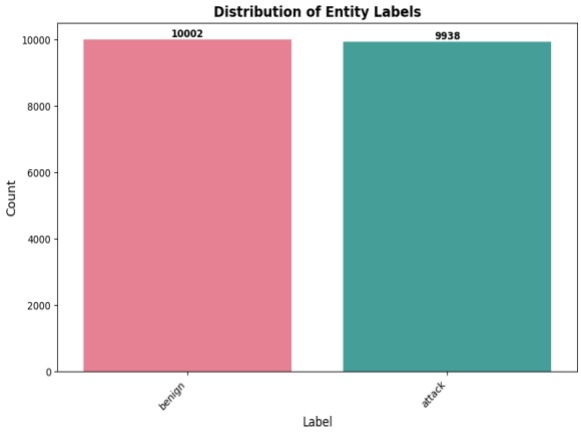

Figure 1. Distribution of Entity Labels

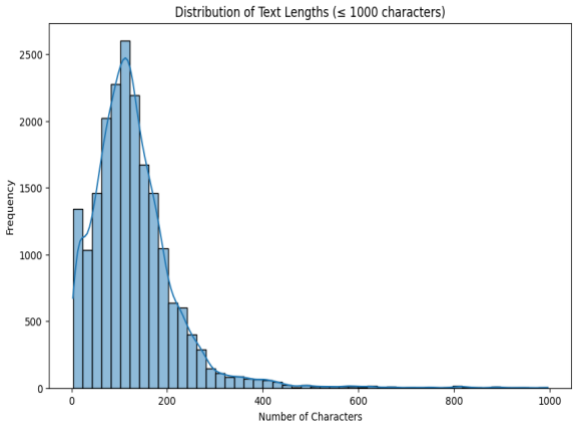

Figure 2. Distribution of Text Lengths

The bar chart in figure 1 indicates that benign and attack samples have almost equal distribution and this means that the dataset is balanced. This balance will prevent the problem of classification bias and will aid more credible training of models. As indicated by the histogram in figure 2, the majority of the text records are relatively short, which corresponds to realistic descriptions of cyber-threats in the real world. The trend curve is smooth, which is another affirmation that the frequency decreases drastically with increase in text length.

Figure 3. Most Frequent Words in Cyber Threat Intelligence Texts

In Figure 3, this word cloud demonstrates the most common words in the cybersecurity text dataset, indicating such important words as attack, malware, system, file, and data. The bigger words depict greater frequency, which means that they are more important in describing cyber-threats. The diagram represents general patterns, actions, and objects associated with bad intent. All in all, it gives a brief summary of the most prevalent vocabulary of cyber-threat intelligence.

3.5 Modelling Approach Using LSTM with Attention

In order to overcome the disadvantages observed in the current research works, the study will use an LSTM with an Attention mechanism to overcome the problem of computational ineffectiveness, and inability to capture long-range dependencies. Although the previous ML and DL models showed high classification rates, they could not tend to give preference to the most valuable textual features in cybersecurity stories, particularly in brief yet informative

threat descriptions. Traditional ML methods require extensive manual feature engineering, and more advanced neural networks in the literature demand high computational needs, which is incompatible with lightweight and real-time settings. The combination of LSTM and Attention mechanism in the model aims to selectively draw attention to the words or patterns associated with threats, eliminate the need to use more complex architectures, and address other limitations such as feature dilution, lower transparency, and slower inference common to most deep learning models.

The LSTM + Attention combination also offers the main benefits to cybersecurity text analysis. The LSTM units are useful in capturing sequential dependency and contextual meaning; thus, the model is able to understand how the threat indicators change throughout a sentence, and the Attention layer adds to this by giving more weight to the most informative parts of the text. This selective weighting does not only enhance the accuracy of the detection, but also helps to get a better explanation of the data as the attention scores also give an explanation of which textual elements contributed to each prediction. Furthermore, since Attention eliminates redundancy in computation by not only concentrating on significant features but also meaningfully, the ensuing architecture is lightweight and applicable to resource-constrained IoT and edge settings. Collectively, these advantages can create a more flexible, effective, and transparent text-based method of cyberattack detection than the traditional ML, single-use DL models, or multi-faceted hybrid systems.

3.6 LSTM–Attention Architecture

The LSTM-Attention architecture applied in the current study is developed to effectively identify sequential patterns in the text of cybersecurity and emphasize the most informative aspects to classify it. Long Short-Term Memory (LSTM) networks can learn time-based relationships in textual data and the model can learn to recognize how the indicators of cyberattack change within a sentence. In contrast to basic RNN models, the gated architecture of LSTM allows to deal with long-range correlations and avoid vanishing gradients, and it is suitable to interpret complex threat descriptions where the key cues can be located at any stage of the text. The words that pertain to malware behaviour, compromise of systems, or attack patterns can be randomly distributed in the sentence in cybersecurity reports, and the LSTM element guarantees that the dispersed features are retained in the memory.

In order to add more information to the model in terms of interpreting subtle cues, we add an Attention layer on top of LSTM. Attention mechanism calculates a relevance score of every word in the input sequence and finds the elements of the text that have the most significant impact on the distinction between attack and benign categories. The Attention layer aims at giving attention to the contextually significant tokens like malware, exploit, phishing, or behavioural verbs that often reflect the malicious intent. This selective concentration also enhances the level of detection, which enables the analysts to visualize what words informed the decision. This clarification is vital in the cybersecurity setting where the transparency and trust towards automated decision-making are crucial.

Their synergistic capabilities of LSTM and Attention directly deal with the weaknesses of previous ML, DL, and ensemble models that were identified in the research gaps. The LSTM-Attention architecture is lightweight and efficient, whereas the traditional models needed massive engineering of features or were computationally expensive and did not suit short and noisy textual threat intelligence. The model eliminates the computation cost of deep stacks of layers, which is replaced by smart prioritization of features, allowing it to be predictive and at the same time incurring low computation cost. This is why it is appropriate to real-time or close-real-time applications of cyber defines, particularly in IoT and edge environments where processing capabilities are limited. In general, the architecture offers a harmonized solution that embodies contextual meaning, and guarantees efficiency in operations to detect cyber threats in textual intelligence.

3.7 Model Evaluation Metrics

To assess the performance of LSTM-Attention model, four important measures are used: Accuracy, Precision, Recall and F1-Score that will guarantee the high accuracy in identifying relevant text of cyberattack. Accuracy is a general score that gives the idea of the correct predictions, whereas Precision is the number of the texts, which are considered attacks and are actually malicious, thus minimizing false alarms. Recall is an evaluation of the model that determines whether it is able to correctly identify actual attack examples to ensure that real threats are not overlooked. F1-Score that is the balance between Precision and Recall is the more reliable measure of the performance under the conditions of cybersecurity, where false positives and false negatives can be very dangerous. A combination of these measures indicates the level of effectiveness of the model in detecting malicious patterns and ensuring that the predictions are consistent and reliable.

4 Design Specification

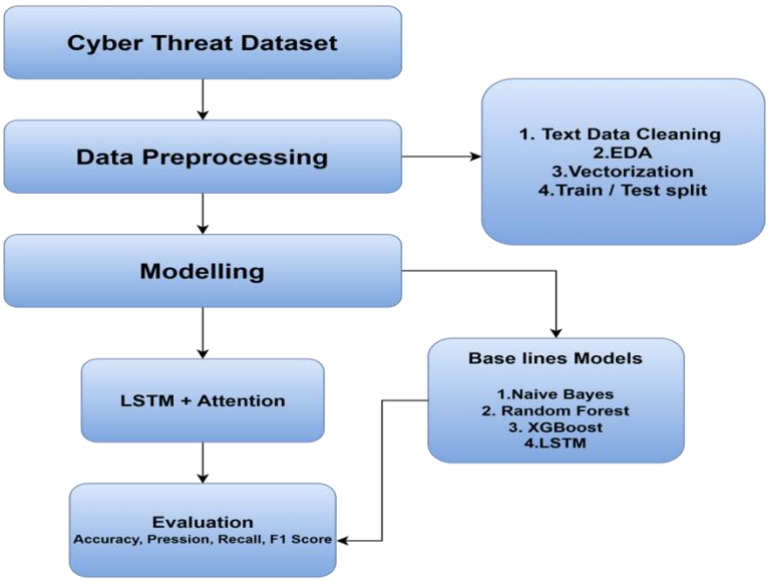

Figure 4. End-to-End Architecture of the Cyber-Threat Classification System

4.1 System Architecture Overview

The architecture of the system presented in figure 4, regarding the cyber-threat text classification framework, is that of a sequential and modular pipeline, as depicted in the design diagram. The Cyber Threat Dataset is the initial stage and the main source of textual threat intelligence that is labelled. This data gets into the Data Preprocessing module where some of the key functions like text cleaning, exploratory data analysis, vectorization and train-test split the raw text into something that can be meaningfully learned using features. After the preprocessing is done, the system goes to the Modelling stage that has two paths one of them being the main advanced model (LSTM + Attention) and the other one being a collection of baseline models (Naive Bayes, Random Forest, and XGBoost and a regular LSTM). These baseline models serve as points of comparison to gauge performance and emphasise the benefits of incorporating attention mechanisms.

Following the modelling phase, the results of the LSTM-Attention model and the baseline models are used as inputs of the Evaluation module, and the predictions are evaluated by traditional performance measures, such as Accuracy, Precision, Recall, and F1-Score. This architecture will also make certain that the system does not only process the text of cyber-threats effectively but also offers a fair comparison of the traditional machine-learning methods, the traditional deep-learning methods, and the improved LSTM-Attention architecture. The workflow is well-organized, logical data flow, and focuses on modularity, and scalability, across the whole pipeline of cyberattack detection.

4.2 Workflow and Pipeline Implementation

The pipeline and workflow application involves the overall processing of cyber-threat text information into a single, modular sequence that guarantees a seamless transit of raw data to the finished model execution. The pipeline starts with loading the cybersecurity dataset, systematic cleaning operations to eliminate noise, normalize texts, and deal with missing labels, encoding, tokenization, and numerical transformation of the text to the machine learning or deep learning models. ML models are feature-scaled, TF-IDF-vectorized, and numerically encoded; deep learning models need padded token sequences, embedded representations and input reshaping to neural-network formats. The pre-processed dataset is divided into stratified training and testing sets after which it can be compared across models. The pipeline will then coordinate the training processes of traditional ML classifiers, deep learning architectures and the LSTM + Attention model with customized hyperparameters, loss functions and optimization strategies. Lastly, the outputs of all the models are then fed into the evaluation phase, in which the accuracy, precision, recall, and F1-score are calculated to determine overall performance and compare the performance of the various modelling methods when used in the same, consistent implementation framework.

5 Implementation

5.1 Machine Learning Models Implementation

For the Naive Bayes model, hyperparameter tuning was performed using a grid search over two key parameters is alpha values 0.1,0.3,0.5,0.7,1.0 and fit prior. These parameters control smoothing strength and whether class priors are learned from the data. After evaluating all combinations through cross-validation, the grid search selected alpha is 0.1 as the optimal smoothing value, with the best estimator adopting this configuration. This indicates that lighter smoothing provided the most accurate probability estimates for distinguishing between benign and attack-related textual patterns as shown in figure 5.

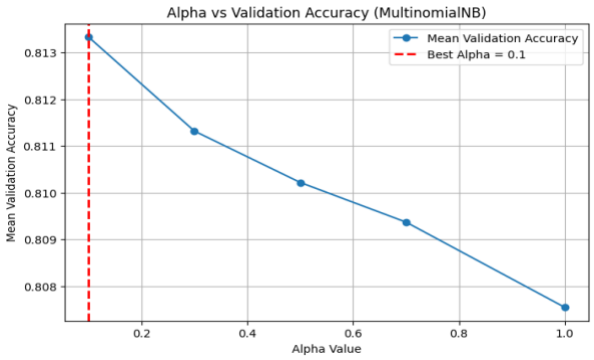

Figure 5. Best Alpha Selection for Naive Bayes

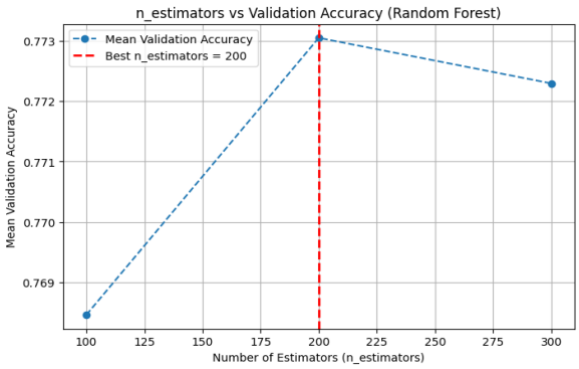

Figure 6. RF hyper parameter tuning with number of estimators

For the Random Forest model, hyperparameter optimization focused solely on varying the number of trees in the ensemble through the number of estimators parameter, tested at values 100,200,300 as shown in figure 6. This parameter determines how many individual decision trees contribute to the final prediction, with larger ensembles typically capturing more variability in the data. After conducting cross-validated search across all candidate values, the tuning process identified number of estimators is 200 as the best setting. This means that an ensemble with a larger number of trees offered the most stable and accurate performance when classifying cyber-threat intelligence text.

For the XGBoost classifier, tuning was performed by adjusting the number of boosting rounds using the number of estimators parameter, explored at values is 50,100,150,200. This parameter directly controls how many sequential trees are built to iteratively correct the errors of previous ones. The grid search evaluated each configuration using cross-validation, and the best estimator was obtained with number of estimators is 200. This shows that the longer the sequence of boosting, the better the model was able to learn complex patterns of textual attacks without the need to add unnecessary complexity.

5.2 Deep Learning Models Implementation

The deep learning model involves an LSTM-based architecture, which is meant to identify sequential patterns in text on cyber-threats, with the first step being preprocessing, which involves cleaning up all the input sentences by lowercasing the text, removing URLs, email addresses and unnecessary characters, to be processed by the neural network. The resulting clean dataset is then tokenized with a vocabulary size of 60,000 words, turned into integer-encoded sequences, and padded to a fixed token length of 500 to make the samples consistent across samples. The computation of class weights is done in order to control the imbalance of classes and facilitate fair learning. The LSTM model is represented in the following way: there is an embedding layer that converts the tokens to dense vectors, and the two subsequent LSTM layers of 256 and 128 units are stacked and each has a dropout layer to avoid overfitting. ReLU activation in a dense layer enhances the learned features and a final sigmoid layer does binary classification. This model is trained with the Adam optimizer, binary cross-entropy loss, and early stopping, and model checkpointing that is used to save the best-performing weights. The trained model is then tested using the test set to produce a score of accuracy, classification metrics, and the confusion matrix to be used as a powerful deep learning baseline in comparison with the traditional machine learning and the LSTM-Attention model.

5.3 LSTM + Attention Implementation

The LSTM + Attention model is executed with the help of structured deep learning pipeline, which starts with the text tokenization and padding, where each sentence is represented as sequences of integers with a vocabulary of 50,000 tokens and padded to a constant length of 200 words. The model architecture begins by embedding a 128-dimensional layer that converts each token into a dense vector representation. This is then followed by an LSTM layer of 128 units that is set up to produce the entire sequence that is needed by the attention mechanism. The Attention layer is customarily followed by calculating the attention weights at every time step, allowing the network to concentrate on the most informative bits of the text. Following attention, the model will have a Dropout layer, a Dense layer with 128 neurons stimulated by ReLU and a second Dropout of to mitigate overfitting. Lastly, a 1-unit, activation sigmoid, Dense output layer is used to achieve binary classification of cyberattack and benign text. The model is trained on the Adam optimizer with a learning rate of 1e-3, the early stopping patience of 10, and the checkpointing to save the weights that perform best. Such a combination of structured encoding, sequential modelling, attention weighting, and optimized training configuration can enable the model to successfully detect the main threat indicators in the text data of cybersecurity.

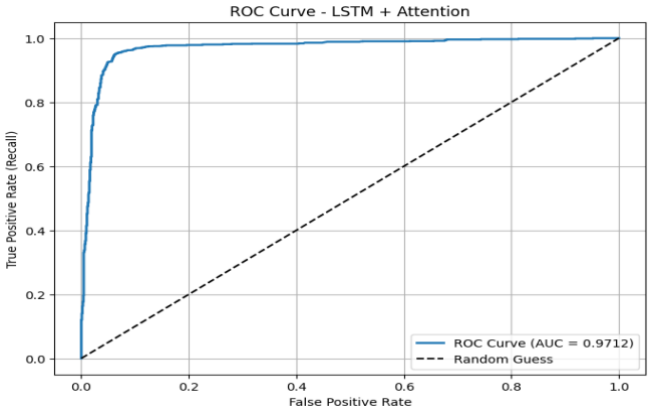

Figure 6.ROC Curve for LSTM + Attention Model

This ROC curve in figure 7 indicates that LSTM + Attention model has a very high true positive rate even when false positive rate is low; which means that it has a strong discriminatory ability. The AUC also attests to the fact that the model works exceptionally well as compared to random guessing.

5.4 Libraries Used

The research makes use of a number of libraries in Python to facilitate the processing of data, model implemented, and evaluation. Data cleaning and numerical operations are processed using pandas, whereas machine learning models, hyperparameter optimization, and evaluation metrics are processed using scikit-learn. Deep learning models such as LSTM and LSTM + Attention are developed with the help of TensorFlow and Keras that offer the layers of the neural-network and tokenization, as well as training tools. XGBoost is applied on gradient-boosted models and Matplotlib can be used to visualize the results of the model, including accuracy plots and confusion matrices. All of these libraries will allow the efficient and full implementation of the system of cyberattacks prediction.

6 Evaluation Results

Table 1. Model Performance Comparison

Model	Accuracy	Precision	Recall	F1 Score
Naive Bayes	82.26%	80.82%	84.30%	82.52%
Random Forest	79.72%	91.24%	65.45%	76.22%
XGBoost	86.56%	91.81%	80.09%	85.55%
LSTM	91.52%	91.82%	91.03%	91.43%
LSTM + Attention	**94.31%**	**93.40%**	**95.29%**	**94.33%**

This section summarizes the performance findings of the entire machine learning, deep learning, and attention-based models to be used in the study before proceeding with the detailed analysis of each of them. These tests indicate the effectiveness of each method in identifying cyberattack-related text in terms of accuracy, precision, recall, and F1-score as illustrated in table1.

6.1 Machine Learning Model Results

Naive Bayes model had an accuracy of 82.26, a precision of 80.82, a recall of 84.30 and an F1-score of 82.52, which implies that the model performs well in all the three metrics. Its relatively high recall implies that it is useful in detecting attack samples, as it should be the case because Naive Bayes can efficiently work with high-dimensional text features. The model is a simple probabilistic model, which can learn sparse representations based on TF-IDF representations and thus is useful in detecting frequent linguistic patterns in cyber-threat text. All in all, Naive Bayes offers good baseline performance at low cost of computation.

The accuracy of the Random Forest classifier was 79.72, and the precision was 91.24 but the recall was much lower (65.45), giving an F1-score of 76.22. This implies that the model is highly conservative and seldom incorrectly identifies benign samples as an attack, though it fails to identify many real cases of attack. The lack of precision-recall balance also indicates that Random Forest, despite its strength with structured tabular data, may not be capable of modelling the semantic complexity of textual threat patterns to the fullest extent compared to models that are designed to operate with sequential or probabilistic representations. However, it is accurate and therefore can be trusted in the event of predictions of an attack.

XGBoost model performed best among all the machine learning models with an accuracy of 86.56, a precision of 91.81, a recall of 80.09, and an F1-score of 85.55. This equal result proves the idea that XGBoost is an efficient tool that determines the majority of attack-related samples with keeping the high degree of prediction reliability. The mechanism of its boosting allows the model to learn minor text characteristics and rectify the previous inaccuracies, which makes it more flexible to the differences in cybersecurity descriptions. Consequently, XGBoost is the strongest among the conventional ML techniques in identifying cyber-attacks in this dataset.

6.2 Deep Learning Model Results

The LSTM deep learning model performed well with an accuracy of 91.52, precision of 91.82, recall of 91.03, and F1-score of 91.43, which means that it is effective in learning sequential and contextual patterns in cyber-threat text. The fact that the model has a close balance between precision and recall demonstrates that the model is capable of detecting malicious samples accurately with low false positives. The LSTM is more robust at capturing subtle semantic connections in language than traditional machine learning techniques, since it uses word embeddings and recurrent layers, which represent a more effective way of capturing finer language nuances in the description of cybersecurity incidents.

6.3 LSTM + Attention Model Results

LSTM + Attention model had an outstanding performance with an accuracy of 94.31, precision of 93.40, recall of 95.29, and F1-score of 94.33, which is the highest performing model in the whole experimental set-up. This great enhancement shows the power of the combination of an attention mechanism with LSTM architecture, which enables the model to selectively attend to the most significant components of every sentence instead of considering all words equally. The description of cyber-threats can have the keywords or context patterns that are critical to the description randomly distributed throughout the text, and attention layer enables the model to give more weight on these indicators, which enhances its potential to differentiate between the cues of maliciousness that are subtle and those that are benign. The high recall value also indicates that the model effectively identifies nearly all the instances related to an attack, which is a crucial aspect of cybersecurity application since the omission of a threat may cause severe outcomes.

Further, the balanced precision and recall relationship point to the fact that the model does not only detect attacks correctly, but also reduces false alarms, which makes it practically reliable in the context of real-world use. Embeddings, LSTM units, and attention weighting together assist the model to learn more semantic relationships, both sequential dependencies and contextual importance of the text in cyber-threats. This renders the LSTM + Attention architecture especially appropriate to intricate, high-dimensional textual intelligence, with meaning and location of words being relevant. Generally, the excellent performance of the model portrays the usefulness of attention-enhanced deep learning methods in the development of scalable and high-accuracy cyberattack prediction systems.

6.4 Findings of the Study and Superiority

The high performance of the LSTM + Attention model is reflected in the results of the study since it achieves high performance in all baselines in terms of accuracy, precision, recall, and F1-score. Although the traditional ML models are efficient in simple text classification, they fail to consider a richer semantic meaning and extended contextual relationships when describing cyber-threats. Even the independent LSTM model, although it can capture sequential information, the importance of all words is equal and cannot focus on the most significant indicators of malicious activity. On the contrary, the LSTM + Attention architecture is superior as it incorporates sequence learning of LSTM with focused attention that highlights keywords, behavioural patterns, and context-specific patterns that are highly related to cyberattacks. This prioritization capability enables the model to notice the presence of subtle threats, minimise the impact of noisy or irrelevant words, and realise more stable generalization. The attention scores are also easier to comprehend which sections of the text contributed to the prediction which can provide better understanding to the cybersecurity analysts. On the whole, the results prove that the addition of attention can greatly contribute to the effectiveness, readability, and predictability of the model, and the LSTM + Attention model is the most stable and context-dependent one used in the current study to detect cyberattacks through text.

7 Conclusion and Future Work

7.1 Conclusion

The paper has shown that machine learning and deep learning combined with advanced attention-based architectures can enhance the accuracy and reliability of cyberattack prediction using textual threat intelligence. Through a systematic comparison of various models such as the Naive Bayes, Random Forest, XGBoost, LSTM and the LSTM + Attention architecture the research identifies the incremental improvement in performance as the models gain more context-awareness and can identify deeper semantic patterns. The major result of the work is the demonstration of the efficiency of the LSTM + Attention model that outperforms all the rest in terms of selective focus on the most important elements of cyber-threat stories, which can offer better detection opportunities, and efficiency in its functioning. Not only does the study provide a strong comparative analysis but also establishes a framework that can be scaled and explained that can be used in the real world where timely and accurate identification of threats is critical.

7.2 Future Work

Future research can be directed towards the development of the proposed system to accommodate additional and more complicated types of cyber-threat information, such as multimodal inputs, such as network logs, URLs, file metadata, and social engineering indicators. The use of transformer-based architectures such as BERT or edge device optimized lightweight versions could further improve contextual understanding without sacrificing efficiency. Also, investigating real-time deployment examples, e.g. its application in intrusion detection systems or threat intelligence platforms, would prove its practical usefulness. The adversarial robustness testing can be also involved in the future research to study the performance of the model under the conditions of obfuscated or manipulated text, and continual learning methods that enable the system to adapt to the new threats independently. All of these directions can be combined to create a more holistic, resilient, and flexible system of cyberattack prediction.

References

Agbaje, M. and Afolabi, O., 2024. Neural Network-Based Cyber-Bullying and Cyber-Aggression Detection Using Twitter (X) Text. Revue d'Intelligence Artificielle, 38(3).

Akar, F., 2024. Performance Analysis of NLP-Based Machine Learning Algorithms in Cyberbullying Detection. Erzincan University Journal of Science and Technology, 17(2), pp.445-459.

Al-Quayed, F., Ahmad, Z. and Humayun, M., 2024. A situation based predictive approach for cybersecurity intrusion detection and prevention using machine learning and deep learning algorithms in wireless sensor networks of industry 4.0. IEEE Access, 12, pp.34800-34819.

Alaca, Y., Celık, Y. and Goel, S., 2023. Anomaly detection in cyber security with graph-based LSTM in log analysis. Chaos Theory and Applications, 5(3), pp.188-197.

Alkhudaydi, O.A., Krichen, M. and Alghamdi, A.D., 2023. A deep learning methodology for predicting cybersecurity attacks on the internet of things. Information, 14(10), p.550.

Alqahtani, A.F. and Ilyas, M., 2024. A machine learning ensemble model for the detection of cyberbullying. arXiv preprint arXiv:2402.12538.

Aliyeva, Ç.O. and Yağanoğlu, M., 2025. Deep learning approach to detect cyberbullying on twitter. Multimedia Tools and Applications, 84(19), pp.20497-20520.

Benavides-Astudillo, E., Fuertes, W., Sanchez-Gordon, S., Nuñez-Agurto, D. and Rodríguez-Galán, G., 2023. A phishing-attack-detection model using natural language processing and deep learning. Applied Sciences, 13(9), p.5275.

Mishra, S., Albarakati, A. and Sharma, S.K., 2022. Cyber threat intelligence for IoT using machine learning. Processes, 10(12), p.2673.

Mubeen, M., Muskan, A., Akram, A., Rashid, J., Alshalali, T.A.N. and Sarwar, N., 2025. Cyberbullying-Related Automated Hate Speech Detection on Social Media Platforms Using Stack Ensemble Classification Method. International Journal of Computational Intelligence Systems, 18(1), pp.1-24.

Nguyen, V. and Kebande, V., 2024. Analysis and Prediction of Cyber-Threats using Machine Learning algorithms.

Okey, O.D., Maidin, S.S., Adasme, P., Rosa, R.L., Saadi, M., Carrillo Melgarejo, D. and Zegarra Rodríguez, D., 2022. BoostedEnML: Efficient technique for detecting cyberattacks in IoT systems using boosted ensemble machine learning. Sensors, 22(19), p.7409.

Purba, M., Paisal, P., Darmo, C.P., Noprisson, H. and Ayumi, V., 2024. MODEL OF INDONESIAN CYBERBULLYING TEXT DETECTION USING MODIFIED LONG SHORT-TERM MEMORY. JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer), 10(1), pp.9-14.

F. Ramoliya, R. Kakkar, R. Gupta, S. Tanwar and S. Agrawal, "SEAM: Deep Learning-based Secure Message Exchange Framework For Autonomous EVs," 2023 IEEE Globecom Workshops (GC Wkshps), Kuala Lumpur, Malaysia, 2023, pp. 80-85, doi: 10.1109/GCWkshps58843.2023.10465168.

Utubor, S., 2023. Improving Detection of Attacks in Cyber-Physical Systems: Applying Gradient Boosting Based Machine Learning Techniques (Doctoral dissertation, The George Washington University).

Vitorino, J., Silva, M., Maia, E. and Praça, I., 2025. Reliable feature selection for adversarially robust cyber-attack detection. Annals of Telecommunications, 80(3), pp.341-355.