A Review of Neural Networks for Enhanced User Entity Behavior Analytics in Cybersecurity: Addressing the Challenge of Vanishing Gradient

[1]Akampurira Paul (Ph.D), [2]Bashir Olaniyi Sadiq Ph.D, [2]Dahiru Buhari Ph.D, [3]Maninti Venkateswarlu Ph.D, [3]Atuhe Aarone, [3]Mugisha Brian

[123]Kampala International University, Uganda

[3]Mbarara University of Science and Technology

DOI: https://doi.org/10.51584/IJRIAS.2024.912016

Received: 27 November 2024; Accepted: 02 December 2024; Published: 03 January 2025

ABSTRACT

In the dynamic realm of cybersecurity, User Entity Behavior Analytics (UEBA) emerges as a pivotal tool, employing advanced data analytics and machine learning to scrutinize user and entity activities, thereby detecting potential insider threats. Recurrent neural networks (RNNs) are particularly notable in this context for their ability to identify complex temporal patterns and strengthen threat detection systems. The vanishing gradient issue, in particular, presents difficulties for efficient model training and convergence, hence the use of RNNs in UEBA is not without its difficulties. This explorative article delves into the nuances of the vanishing gradient issue within RNN architectures in the context of UEBA. By dissecting the challenge and exploring potential solutions, we aim to provide readers with a comprehensive understanding of the complexities involved and pave the way for future research directions in optimizing RNNs for enhanced cybersecurity applications. Our analysis reveals important gaps in the application of neural networks for insider threat prediction and the advancement of behavior analytics systems. We give important insights into the challenges of handling the vanishing gradient, and we conclude by recommending neural network optimization methods for behavior analytics and cybersecurity to scholars, practitioners, and organizations.

Keywords: Recurrent Neural Networks (RNNs), Cybersecurity, Insider Threat Detection, Vanishing Gradient Problem, Intrusion detection systems, anomaly detection, Machine Learning, Data Analytics.

INTRODUCTION

Cybercrime costs have recently dramatically surged , and predictions indicate that there has been a significant increase from $3 trillion in 2015 to over $6 trillion in 2021, its projected to rise to $10.5 trillion by 2025 and an alarming $23.84 trillion by 2027 (Riek & Böhme, 2018; Hawdon, 2021). Similarly, insider threats in organizations and businesses have increased by 47% in the last two years. This makes up 60% of data breaches, costing an average of $15.38 million each event, with expenses rising by 40% in the last four years according to the world economic summit reports. (Jurgens, 2023; Fleck, 2024).

Recently, cutting-edge machine learning methods like Recurrent Neural Networks (RNNs) have been utilized to examine user and entity activity, User Entity Behavior Analytics (UEBA) has become essential in response to this growing threat landscape (Shashanka et al., 2016; Salitin et al., 2023). These systems are made to track and examine how users and other entities behave within a network, looking for trends that can point to possible security risks like insider assaults. UEBA uses machine learning algorithms to identify unusual activity that could indicate an insider threat or other sophisticated attacks, in contrast to standard cybersecurity solutions that concentrate on known threats (Olaniyan et al., 2023).

RNNs are especially well-suited for simulating user behavior over time because of their capacity to process sequential input. RNNs are excellent at identifying subtle, changing risks because they can identify patterns in action or event sequences that static models could overlook. However, there are obstacles to RNN deployment in UEBA, particularly the vanishing gradient issue, which calls for thorough investigation to clarify its complexities and possible solutions (Yuhuang, 2019). Gradients attenuate exponentially during backward propagation across network layers, which makes it difficult to implement efficient weight updates and jeopardizes model convergence. This problem is particularly evident in RNNs (Roodschild, 2020; Hu, 2018).

This drastically limits RNNs' performance in jobs that require extended context, like UEBA, by making it difficult to understand long-term dependencies. The disappearing gradient issue can lead to severe performance loss in cybersecurity, where the capacity to record and examine lengthy user activity sequences is crucial for precise threat identification. The vanishing gradient problem is made worse by a number of reasons, including the saturation tendencies of traditional activation functions, deeper network depth, and the rapid reduction of gradients when modeling complex long-term dependencies in RNNs (Li, 2017). Resolving this issue is essential for deep learning architectures, especially in cybersecurity's UEBA, where the vanishing gradient issue interferes with weight updates and model convergence (Hu, 2018; Roodschild, 2020).

In order to improve RNN stability and performance, this review suggests an organized framework that integrates bidirectional designs and innovations such as LSTM networks, GRUs, ESNs, and Transformers, as well as assessing optimization paradigms (Shrestha, 2023; Talaei Khoei, 2023). The work demonstrates how these optimization methods and architectural developments effectively address the vanishing gradient issue and improve RNN performance in sequential data modeling applications.

To fully utilize deep learning architectures, particularly in domains like cybersecurity's User Entity Behavior Analytics (UEBA), this difficulty must be resolved. The vanishing gradient issue has several consequences, including impaired neural network performance, hampered model convergence, and compromised weight updates. Remedial actions are necessary to fully realize the potential of deep learning paradigms in cybersecurity domains, particularly UEBA, as these obstacles lead to inferior performance metrics and increased computing complexity. The research community has aggressively started developing novel approaches and solutions to mitigate the vanishing gradient dilemma and improve RNN performance in response to these demands (Shrestha A, 2023; Talaei Khoei T, 2023).

In its 2023 report, the World Economic Forum highlights the importance of improving communication between company executives and cybersecurity, defining organizational responsibilities, and cultivating a strong security culture. It emphasizes the necessity of implementing diversity and training programs to close the cybersecurity talent gap. The paper also identifies developing factors that will shape the global cyber scene in 2023, including as geopolitics, new technologies, changing threats, and legislative changes (Jurgens P. D., 2023). The restrictions caused by the vanishing gradient problem must be addressed since RNNs are essential to cybersecurity, especially in UEBA systems.

To improve the efficacy of RNNs in cybersecurity applications, this review attempts to methodically assess the body of research on RNN optimization strategies that can lessen this problem. Through the identification and evaluation of the most promising methods, this study will offer practitioners and researchers useful information for creating reliable and effective RNN models for UEBA.

Therefore, as cybersecurity threats grow increasingly complex, and traditional models fall short. And RNNs' ability to monitor user behavior over extended periods as an essential capability for identifying subtle and evolving risks, is also affected by vanishing gradients issues. Thus, resolving the this issue is critical for enhancing RNN performance in UEBA systems. In this study therefore, we evaluated the existing literature on RNN optimization techniques, focusing on architectural advancements, optimization strategies, and algorithms that can improve the stability and performance of RNNs in sequential data modeling tasks relevant to intrusion detection and insider threats. By addressing this issue, our study seeks to provide valuable insights for researchers and practitioners, ultimately contributing to the development of more robust cybersecurity solutions capable of adapting to an ever-evolving threat landscape.

FLOWCHART FOR REVIEW

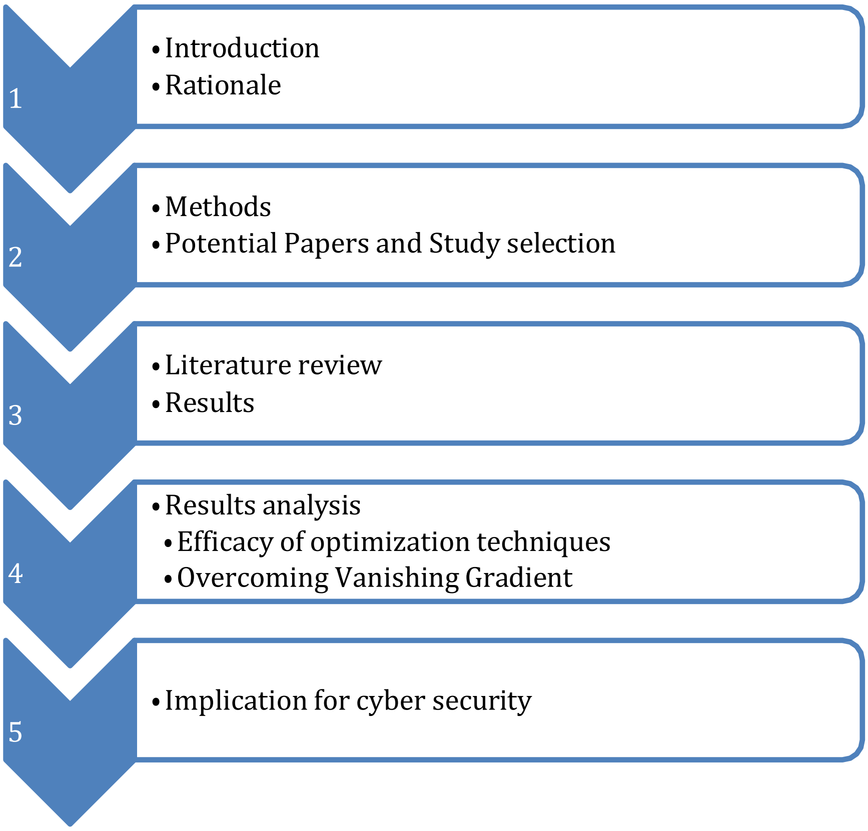
1
• Introduction
• Rationale
2
• Methods
• Potential Papers and Study selection
3
• Literature review
• Results
4
• Results analysis
• Efficacy of optimization techniques
• Overcoming Vanishing Gradient
5
• Implication for cyber security

Figure 1: Flowchart for Review

From the chart above, we note that this review process was designed to systematically identify, evaluate, and analyze research literature focused on the optimization of Recurrent Neural Networks (RNNs) in cybersecurity, particularly addressing the vanishing gradient problem.

MATERIALS AND METHODS

The section is divided into key phases: Identification of Potential Papers, Study Selection, and Data Extraction and Analysis. Each phase employs a rigorous methodology to ensure the inclusion of high-quality and pertinent studies, providing a comprehensive overview of advancements in RNN optimization techniques and their applications in User Entity Behavior Analytics (UEBA).

Identification of Potential Papers

To ensure a comprehensive collection of relevant literature, we conducted our search across multiple well-established electronic databases and digital libraries. The databases selected for this review include: IEEE Xplore: Known for its extensive collection of technical literature in engineering, computer science, and related fields, IEEE Xplore was a primary source for research papers on RNNs and cybersecurity: ACM Digital Library: This digital library houses a vast array of computing and information technology resources, providing access to studies focused on RNN architectures and their application in cybersecurity: SpringerLink: Springer’s platform was utilized to find academic papers and book chapters that discuss advanced machine learning techniques, including RNNs, LSTMs, and GRUs: ScienceDirect: As a leading full-text scientific database, ScienceDirect was instrumental in sourcing literature on RNN optimization techniques and their application in real-world cybersecurity scenarios: and Google Scholar: As an aggregator of scholarly articles, Google Scholar was used to supplement the search by identifying additional relevant studies that might not be indexed in the aforementioned databases. The search procedure was carefully crafted to ensure that only the most relevant and high-quality papers were selected. The following steps were involved in the search procedure: Keywords: The search was conducted using a combination of keywords that reflect the focus of the study. These included terms such as “Recurrent Neural Networks,” “Vanishing Gradient,” “LSTM,” “GRU,” “Attention Mechanisms,” “UEBA,” “Cybersecurity,” “Optimization Techniques,” and “Insider Threat Detection.”: Search Strings: The search was limited to peer-reviewed journal articles and conference papers published within the last decade (2014-2024). This ensured that the review focused on contemporary advancements and methodologies. Additional filters were applied to exclude non-English language papers, ensuring accessibility and consistency.

Study Selection

The following inclusion criteria were established to guide the selection of studies for the review: Only studies that explicitly focused on the optimization of RNN architectures, particularly in addressing the vanishing gradient problem, were included. Also, papers that applied RNNs to cybersecurity tasks, such as UEBA, anomaly detection, and insider threat detection, were prioritized. Additionally, to ensure academic rigor and credibility, only articles published in peer-reviewed journals or reputable conference proceedings were considered. And finally, preference was given to studies published within the last decade to capture the latest advancements and trends.

The exclusion criteria were designed to filter out studies that did not align with the review's objectives: Studies that did not reflect recent advancements in RNN technology or addressed issues already considered resolved were excluded. Also, papers that discussed RNN optimization in contexts outside of cybersecurity, such as natural language processing or general machine learning tasks, were excluded unless they provided transferable insights. Additionally, studies that did not address the vanishing gradient problem, or that focused on other optimization challenges not directly related to RNNs in cybersecurity, were excluded from the review. Lastly, papers published in non-peer-reviewed sources, such as white papers, preprints, or blog posts, were excluded to maintain academic integrity.

RESULTS

In this phase, we conducted an in-depth analysis of the selected studies to gain insights into the methods and strategies employed to optimize Recurrent Neural Networks (RNNs) for cybersecurity applications. The review is details key architectures including Traditional RNNs, the vanishing gradient problem, Schemes and optimization methods, and design strategies involving LSTM, GRUs, ECN and attention mechanisms.

Traditional RNNs and Vanishing Gradient

According to Das, 2023, Recurrent Neural Networks (RNNs) are a specialized type of neural network particularly suited for processing sequential data, which can represent temporal patterns or ordered events. Due to their ability to maintain memory of past inputs through hidden states, RNNs are effective in identifying patterns in time series data, such as user behavior in cybersecurity systems. This capability is essential when tracking anomalous behavior over time, making RNNs valuable in detecting insider threats and suspicious activity in Intrusion Detection Systems (IDS). (Shiri, 2023) .

In the context of insider threats and IDS, RNNs can capture the sequential nature of user actions while identifying deviations from normal behavior that may signal malicious intent and can respectively be used to predict and detect anomalous network traffic patterns, intrusions, or other forms of suspicious activity by learning from historical data. An RNN processes a sequence of inputs ($x = (x_1, x_2, ..., x_T)$ and produces a sequence of outputs (y = (y_1, y_2, ….., y_T)). At each time step (t), the hidden state (h_t) is updated based on the previous hidden state (h_{t-1}) and the current input (x_t). This can be represented as:

$$h_t = \sigma\ (W_{xh}x_t + W_{hh}h_{t-1} + b_h)$$

where;

h_t is a hidden state at time step tt.

x_t is input at time step t.

W_{xh} is the weight matrix for the input to hidden state.

W_{hh} is the weight matrix for the hidden state to hidden state.

b_h is the bias term for the hidden state.

σ is an activation function (commonly tanh or ReLU)

And the output function

$$y_t = \phi(W_{hy}h_t + b_y)$$

where;

y_t is output at time step t.

W_{hy} is the weight matrix for the hidden state to output.

b_y is the bias term for the output.

And ϕ is an activation function (commonly softmax for classification tasks).

The hidden state plays a crucial role in retaining information from previous steps (e.g., past user behavior in a system or network activity logs). In insider threat detection, the hidden state allows the RNN to model patterns over time, such as a user's login activities or file access patterns. This enables the system to detect subtle deviations indicative of insider threats (Wei Hong, 2023). Weights in RNNs are shared among various time steps. Every time step, the input is processed, and the hidden state is updated using the same set of weights and biases. RNNs can capture temporal dependencies in sequential data and retain the knowledge of prior time steps because of this weight sharing. In IDS, this output represents a probability score indicating whether the current network activity is normal or anomalous. Therefore, by continuously updating the hidden state and producing outputs over a sequence of network packets, the RNN can provide real-time predictions about potential intrusions (Das, 2023).

Despite their strengths, RNNs face a significant challenge known as the vanishing gradient problem. This issue arises during the training of the network, particularly when using gradient-based optimization methods like backpropagation through time (BPTT). The problem occurs because the gradients of the loss function with respect to the weights can become exceedingly small, effectively vanishing as they are propagated back through many layers (or time steps). When the gradients are too small, the weights are updated very slowly, and the network learns very little with each training iteration (Hu, 2018). This is particularly problematic for long sequences, where the influence of an input on the hidden state and output diminishes exponentially as it moves further back in time. Consequently, the network struggles to learn long-term dependencies, which are often crucial for tasks involving sequential data.

Many variants of RNNs have been devised including bidirectional RNNs which processes input sequences in both forward and backward directions. It consists of two separate hidden layers for each time step, one processing the sequence from the beginning to the end (forward), and the other processing it from the end to the beginning (backward). BRNNs capture information from past and future contexts simultaneously, making them especially useful for tasks where context from both directions is important (Das, 2023). And, Deep RNN which make use of multiple hidden layers in the recurrent network architecture. Each layer processes the input sequence, and the output of one layer serves as the input to the next layer. Deep RNNs can potentially capture more complex hierarchical features in sequential data. They can learn abstract representations at different levels of abstraction (Li, 2019). Both Bidirectional RNNs and Deep RNNs can be used independently, but they can also be combined to form Bidirectional Deep RNNs, which leverage the advantages of both bidirectional processing and deep architectures (Miftahutdinov, 2019).

The primary advantage of using RNNs in cybersecurity is their ability to handle sequential data and capture temporal dependencies, which are crucial for detecting sophisticated attacks. However, training RNNs can be challenging due to issues like vanishing and exploding gradients. Techniques such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) have been developed to mitigate these issues and enhance the performance of RNNs in practical applications. (Matilda Rhode, 2018).

The vanishing and exploding Gradients' phenomena are common in RNNs and occur due to the multiplicative nature of gradients, which can exponentially decrease or increase with the number of layers. This makes it challenging to capture long-term dependencies. In the essence of the vanishing gradient problem, the gradients of the loss function with respect to the weights in the early layers can become very small, effectively preventing these weights from being updated during training (Hu Z, 2021; Hu, 2018; Roodschild M, 2020; Takudzwa Fadziso, 2020).

Consequently, this phenomenon limits the network's ability to capture long-term dependencies in sequential data and impedes its learning capacity (Das, 2023). The exploding gradient problem can be managed using gradient clipping, a technique that caps the maximum value of the gradient to control this issue in practice (Ribeiro, 2019).

Addressing Vanishing/exploding gradient problem:

A common deterrent method is gradient clipping technique for handling the exploding gradient problem sometimes encountered when performing backpropagation in RNNs. By capping the maximum value for the gradient, this phenomenon of the exploding gradient is controlled in practice (Pascanu R. M., 2012; Pascanu R. M., 2013). Also, using different activation functions, such as ReLU (rectified linear unit), that do not saturate or have zero gradients for large inputs. ReLU is defined as $f(x)=max(0,x)$ and has a constant gradient of 1 for positive inputs and 0 for negative inputs. This helps to avoid the multiplication of small gradients between layers and allows the gradients to flow more easily through the network (Agarap, 2021).

Additionally, employing different weight initialization schemes, such as He initialization, scales the weights according to the number of input and output units of each layer. This helps to avoid the problem of exploding or vanishing gradients by maintaining consistent variance of the weights and activations throughout the network. Also, using optimization algorithms like Adam, which adapt the learning rate for each weight based on the history of the gradients, also helps. This approach overcomes the challenge of selecting a suitable learning rate for the entire network and prevents issues of slow convergence or divergence due to inappropriate learning rates (Ribeiro, 2019)

Moreover, applying different network architectures, such as ResNet (residual network), that add skip connections or shortcuts between layers. This helps to avoid the problem of vanishing gradients by allowing the network to learn residual functions instead of direct mappings. This also helps to improve the performance and generalization of the network by enabling it to learn from both low-level and high-level features (Lu W, 2021; F. Meng, 2021)

However, these solutions also have some limitations, for example, the ReLU activation function can cause the problem of dying neurons, which means that some neurons can become inactive and stop learning if their inputs are always negative. This can reduce the capacity and diversity of the network and lead to poor performance. To overcome this problem, some variants of ReLU, such as Leaky ReLU or ELU (exponential linear unit), have been proposed that have non-zero gradients for negative inputs (Agarap, 2021)

Also, He initialization and Adam optimization algorithms can cause the problem of overfitting, which means that the network can memorize the training data and fail to generalize to new or unseen data. This can reduce the robustness and accuracy of the network and lead to poor performance. To overcome this problem, some regularization techniques, such as dropout or batch normalization, have been proposed that reduce the complexity and dependency of the network and improve its generalization ability (Kingma, 2014)

Moreover, ResNet architecture can cause the problem of computational complexity, which means that the network can have a large number of parameters and layers that require more memory and processing power to train and test. This can limit the scalability and efficiency of the network and lead to high costs and time. To overcome this problem, some compression techniques, such as pruning or quantization, have been proposed that reduce the size and redundancy of the network and improve its speed and performance.

RNN Optimization architectures

RNN optimization strategies, such as Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRUs), Echo State Networks (ESNs), and attention mechanisms, are crucial for enhancing insider threat detection and user behavior analytics (UEBA).

Long Short-Term Memory (LSTM)

Houdt, (2020), Reviewed the Long Short-Term Memory by Hochreiter and Schmidhuber (1997) which introduced the LSTM architecture, which is suitable for modeling sequential data and can be tailored for user behavior analytics. The study concluded that LSTM showed promising results in capturing temporal dependencies, making it suitable for modeling user behavior in cyber threat detection. The author recommended exploring different variants of LSTM, such as stacked LSTM or bidirectional LSTM, to enhance the modeling capabilities for user behavior analytics.

Also, study by (Zhaoyang Niu, 2023), about the gated recurrent unit architecture, which is another variant of recurrent neural networks suitable for sequential data analysis. In the study, GRU demonstrated comparable performance to LSTM while having a simpler structure, making it a potential choice for user behavior analytics. It was recommended that further studies can progress to comparing the performance of GRU with LSTM and exploring hybrid models that combine the strengths of both architectures.

Moreover, (Bin Sarhan & Altwaijry, 2023), in a study on LSTM-based User Behavior Analytics for Cyber Threats about user behavior analytics in the context of cyber threat detection effectively captured sequential patterns in user behaviors and detected anomalous activities indicative of cyber threats. The authors recommended exploring the integration of attention mechanisms and reinforcement learning techniques to enhance the model's performance in detecting sophisticated cyber threats.

In addition, (Mahmoud Abbasi, 2021) studied anomaly detection in network traffic based on wavelet transform and LSTM neural networks, and proposed an anomaly detection framework using wavelet transform and LSTM neural networks for network traffic analysis. Their model achieved improved performance in detecting anomalous user behaviors in network traffic compared to traditional methods. The authors recommended conducting comparative evaluations of LSTM-based models with other deep learning architectures for user behavior analytics in cyber threat detection

(Kiran Kumar, 2023) in a study on Optimizing LSTM and Bi-LSTM models, offers valuable insights into LSTM in crop yield prediction and achieved percentage reductions in error with the Bi-LSTM model compared to traditional ML techniques, notably 94%, 72%, and 71% for wheat, groundnut, and barley yield predictions, respectively. While the research effectively highlights the superior performance of deep learning over conventional methods, it lacks depth in explaining the underlying reasons for Bi-LSTM's efficacy and falls short of providing model interpretability. Additionally, there's a need for a more comprehensive discussion on dataset characteristics, external validation methods, and potential limitations. It is important to incorporate feature importance analyses for model interpretability, detailing the dataset's nuances, and employing external validation techniques to bolster the study's credibility and applicability.

(Wang A. M., 2023) in a study that focuses on detecting malicious cyber-attacks using a Deep Learning approach, specifically employing Bi-LSTM models, utilized the UGR’16 dataset and presented Bi-LSTM with an autoencoder and Bi-LSTM without an autoencoder. Notably, the Bi-LSTM model without the autoencoder yielded a remarkable accuracy rate of 99%, outperforming its counterpart, which achieved an accuracy of 93%. This stark difference underscores the effectiveness of the Bi-LSTM model in detecting malicious activities in cybersecurity without the need for additional feature engineering via an autoencoder. Emphasis is made on data preprocessing, exploratory data analysis, and feature extraction using autoencoders.

(Zhao, 2021) compare the performance of an LSTM-based anomaly detection model (BEM) with three baseline models (one-class SVM, GMM, and PCA) on two datasets of security logs (Knet2016 and R6.2). The performance is measured by the AUC, which indicates the trade-off between true positives and false positives1. The paper achieves good results where on the Knet2016 dataset, BEM achieves the highest AUC of 0.999 at the log-line level, and 0.978 at the user-day-max level. The baseline models have AUCs ranging from 0.500 to 0.940 at the user-day-max level. With the R6.2 dataset, BEM achieves the highest AUC of 0.999 at the log-line level and 0.992 at the user-day-max level. The baseline models have AUCs ranging from 0.500 to 0.998 at the user-day-max level.

A study by (Bakhsh, 2023) on enhancing IoT Network Security through Deep Learning-Powered Intrusion Detection System, proposes the use of Deep Learning (DL) techniques, specifically Feed Forward Neural Networks (FFNN), Long Short-Term Memory (LSTM), and Random Neural Networks (RandNN), for enhancing IoT network security. The research demonstrates impressive performance metrics, with the FFNN model achieving an accuracy of 99.93%, the LSTM model achieving 99.85%, and the RandNN model reaching 96.42% in detecting intrusions using the CIC-IoT22 dataset. The authors emphasize the significance of parameter optimization for each DL model, highlighting the adaptability and potential benefits of each technique. They underscore that while FFNN excels in handling complex IoT network traffic patterns, LSTM effectively captures long-term dependencies, and RandNN leverages its random connections for adaptive learning.

Gated Recurrent Unit (GRU):

Gates: Recurrent neural networks (RNNs) are robust models for sequential data modeling in machine learning and artificial intelligence, but they are susceptible to suffer from the vanishing gradient problem. The problem makes it difficult for the model to learn long-term dependencies. One studied method to address this problem is to use Gated Units. These are employed in variants of Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) (Shiri, 2023; Hu Z, 2021; Hu, 2018).

According to study by (Wang A. M., 2023), gated units can control the flow of information and gradients through the network. Gated units have an internal state that can store and update information over time, and a set of gates that can regulate how much information is added or removed from the state. By using gates, RNNs can learn to preserve relevant information and forget irrelevant information, thus alleviating the vanishing gradient problem and improving the performance on tasks that require long-term memory (Wang A. M., 2023; Wang W. R., 2017; W. Liu, 2017).

A study on personalized session-based recommendations with recurrent neural networks" by Hidasi et al. (2015), focused on session-based recommendations using the GRU architecture, where user behavior sequences are modeled. The designed GRU4Rec model demonstrated effective personalized recommendations based on user behavior analytics, showing the potential for cyber threat detection. The authors recommended evaluating the performance of GRU4Rec in the context of cyber threat detection and exploring adaptation strategies for this specific domain.

A study by (Kim, 2022) on threat classification model for security information event management focusing on model efficiency, designed a GRU-based deep learning model specifically tailored for user behavior analytics in cyber threat detection. The model demonstrated competitive performance in identifying malicious behaviors and detecting cyber threats in real time. The authors recommended incorporating temporal attention mechanisms and ensembling multiple GRU models to improve the accuracy and robustness of the model.

A conceptual framework for insider threat identification based on insider behaviors is proposed in a 2019 paper by (Al-Mhiqani, 2020). Furthermore, a strategy for improving insider threat detection the gated recurrent unit (GRU) neural network is investigated. Using simple-structure GRUs, which save training time and additional computing resources, GRUs can capture long-term temporal dependencies on the sequence of user actions well considering the hidden units that GRUs use to record temporal behavior patterns. This is similar to the long short-term memory (LSTM) approach.

In the paper of (Alaca, 2023) on graph-based log anomaly detection, it was discussed how to effectively identify cyberattacks by analyzing log data using the GLSTM framework. With the help of LSTM methods for training and Node2Vec for graph transformation, the suggested GLSTM model produced an impressive 97.01% accuracy rate on a variety of datasets, including HDFS, BGL, and IMDB. Even while the study performed better than previous approaches, particularly when processing textual and numerical log data it also brought attention to some drawbacks, including restrictions related to computing scalability and data variety. The authors proposed reducing computing overheads, exploring alternate deep learning approaches, and improving the model's adaptability across different datasets. Improving the study's practical relevance and streamlining feature representation might further increase its real-world application and encourage wider research uptake.

(Kim, 2022) attempts to develop an AI-based threat classification model. The study compares distinct deep learning models for threat classification within Security Information Event Management (SIEM) systems. Notably, the findings highlight CNN-static(1D) as the optimal model, characterized by a commendable accuracy rate of 94.07% recall coupled with expedited learning and classification durations of 3346 seconds and 5.1 seconds, respectively. Consequently, the research advocates for the integration of CNN-static(1D) within SIEM frameworks, envisioning enhanced operational efficiency and heightened responsiveness to emerging cyber threats, thereby alleviating the burden on security analysts.

(F. Meng, 2021) in a study about GRUs and multi-autoencoder-based insider threat detection for cyber Security, proposed a method based on GRUs and had a recall of 98.7% on CERT v6.2 and 99.1% on CERT v4.2, while the best baseline method had a recall of 82.3% on CERT v6.2 and 85.1% on CERT v4.2. The study presented a pioneering approach for detecting insider threats in cybersecurity using GRU and multi-autoencoder techniques, showcasing superior performance on the CERT dataset. However, the study did not elucidate the reasons behind detected anomalies, emphasizing the need for incorporating explainable artificial intelligence (XAI) methods to clarify insights for security analysts. The comparative analysis was limited to outdated deep learning models, suggesting that forthcoming research should contrast the proposed method with contemporary models like variational autoencoders, generative adversarial networks, and transformers to ensure robustness and relevance.

Echo State Networks (ESN):

(Sun, 2022) explored the designs and applications of Echo State Networks (ESNs). They highlighted ESNs' emergence as a simpler alternative to gradient descent-based Recurrent Neural Networks (RNNs), offering practicality, conceptual simplicity, and ease of implementation. Despite their apparent simplicity, successful ESN applications require experience due to inherent complexities. The authors categorized ESN-based methods into basic ESNs, Deep ESNs, and combinations, analysing them from theoretical, design, and application perspectives. They discussed the progress brought by abundant works and the potential of the recently introduced Deep ESN model to integrate deep learning and ESNs, while also addressing challenges and proposing future research directions.

Moreover a study by (Li, 2022) on Echo State Networks (ESNs) which are known for their reservoir computing capabilities, showed that ESNs demonstrated promising results in modeling complex dynamic systems and could be adapted for user behavior analytics in cyber threat detection. They have been successfully applied in various domains, including modeling complex dynamic systems. The fact that ESNs have shown promising results in this area suggests that they may be suitable for modeling user behavior patterns as well.

Another study on echo state networks for user behavior analytics in cyber threat detection by (Singh, 2023) for user behavior analytics with a focus on cyber threat detection. The ESN effectively captures the dynamic patterns of user behaviors and shows promising results in detecting cyber threats. authors recommend exploring the integration of graph-based modeling techniques and transfer learning approaches to enhance the ESN model's performance in complex cyber threat scenarios.

Transformers. The Attention model:

"Attention Is All You Need" by (Vaswani, 2017) introduced the Transformer architecture, which utilizes self-attention mechanisms for sequence modeling tasks. Transformers have shown state-of-the-art performance in various natural language processing tasks and can be tailored for user behavior analytics. The authors recommend exploring the adaptation of Transformers to model user behavior sequences and incorporating domain-specific features to enhance cyber threat detection capabilities.

According to (Vaswani, 2017), the model allows an RNN to pay attention to specific parts of the input that are considered as being important, which improves the performance of the resulting model in practice. By noting $\alpha^{t,t'}$ the amount of attention that the output y^t should pay to the activation $a_{t'}$ and c^t the context at time t· we have:

$$\sum_{t'} \alpha^{(t,t')} = 1$$

Attention weight: The amount of attention that the output y^t should pay to the activation $a^{t\prime}$ is given by $\alpha^{t,t\prime}$ computed as follows:

$$\alpha^{(t,t')} = \frac{\exp\,(e^{(t,t')})}{\sum_{t''=1}^{T_x} \exp\,(e^{(t,t'')})}$$

The fundamental idea behind attention is to enable the model to focus on specific parts of the input sequence, giving more weight to relevant information while disregarding less important details. This attention mechanism has proven to be particularly effective in tasks involving sequential data, such as natural language processing and machine translation. The attention model thus enables the network to selectively attend to parts of the input sequence that are considered important for the current context, allowing the model to capture long-range dependencies and improve its overall performance. This is especially beneficial in tasks where certain parts of the input sequence contribute more to the output than others (Rafiq, 2023; Andrea Galassi, 2021)

A study by (Demertzis, 2023) demonstrates promising advancements in anomaly detection, outperforming traditional methods by significant margins, as evidenced by the performance metrics: CM-DANA achieved 92% accuracy, 88% precision, 92% recall, and an F1 score of 90%. While its dynamic attention mechanism and cross-modal learning contribute to its robustness, there are concerns about its computational overhead, as indicated by a processing time of 315.2 seconds. Recommendations for future iterations include optimizing computational efficiency, delving deeper into model interpretability, and expanding testing across diverse and larger datasets to ensure scalability and adaptability in real-world scenarios. We affirm that the Transformer architecture, through its attention mechanisms, provides a robust framework for capturing intricate relationships within sequential data. By leveraging scaled dot-product attention, multi-head mechanisms, and positional encodings, Transformers excel in tasks requiring contextual understanding and feature extraction. This mathematical foundation underscores the versatility and efficacy of attention models, propelling advancements across various domains and applications.

ANALYSIS OF RESULTS

The landscape of deep learning, particularly within cybersecurity and other specialized domains, has seen a plethora of advancements over recent years. This discussion delves into the empirical results of several studies, shedding light on the efficacy of proposed optimizations, comparing metrics with baseline models, and drawing insights from observed trends.

Empirical Results and Efficacy of Proposed Optimizations:

The review revealed that several optimization methods significantly enhance the performance of RNNs in in insider threats analysis, intrusion detection and prevention systems and UEBA. Techniques like gradient clipping, learning rate adjustments, and advanced initialization methods have shown promising results in maintaining stability and efficiency in sequential data modeling tasks. Additionally, hybrid models that integrate RNNs with other neural network architectures were found to improve model robustness, especially in handling diverse cybersecurity threats. The effectiveness of these methods is evident in their ability to reduce training times, improve convergence rates, and enhance the detection of anomalous user behavior, which is critical in UEBA systems.

The review also highlights a clear trend towards the adoption of advanced RNN architectures, such as LSTMs and GRUs, in addressing the vanishing gradient problem. The introduction of attention mechanisms has further enhanced the ability of these models to process and analyze sequential data with greater accuracy and efficiency. A summary of some studies are highlighted in the table1 below;

SUMMARY OF RESULTS AND CURRENT MODELS' PERFORMANCE

Table 1: Summary of Current models' performance

Year and year	Title	Dataset	Model	Accuracy	Precision	Recall	Specificity	Sensitivity	ROC-AUC	F1 score	Loss
Nasir, et al. (2021)	Behavioral Based Insider Threat Detection Using Deep Learning	CMU-CERT r4.2	LSTM	90.60	0.97					94	0.31
Vinayakumar, (2020)	Evaluation of RNNs and Variants for IDS	KDDCup'99	RNN, LSTM, GRU	96.98	1.00	0.999			0.999	0.99	0.01
		UNSW-NB15		89.99	0.889	0.973				0.929	0.22
Mariam Ibrahim, (2023)	Modeling an intrusion detection using recurrent neural networks	NSL-KDD	LSTM-RNN	88.4%	0.866			0.885			
Bushra Bin Sahan, (2022)	Insider threat using Machine learning	CERT r4.2	SVM (No SMOTE)	100%	1.00	1.00		1.00			
Fatima Ezzahra, (2021)	Intrusion detection systems using LSTM	KDD-99	LSTM-PCA	99.44%	1.00		0.99		0.99		
Duc C. Le & Nur Zincir-Heywood (2020)	Exploring anomalous behaviour detection and classification for insider threat identification	CERT r4.2	Random Forest	98.33%	0.91			0.70			
		CERT r5.2	Random Forest	99.04%	0.96			0.76			
		CERT r6.2	Random Forest	95.94%	0.20			0.60			
Kiran Kumar (2023)	Optimizing LSTM and Bi-LSTM models	(IMD)		94%							
Zhao, Z. (2021)	A LSTM-Based Anomaly Detection Model for Log Analysis	Knet2016 R6.2 dataset	SVM, GMM and PCA						0.978 0.992		
Zhanquan Wang (2023)	Deep Learning for Malicious Cyber-Attack Detection	UGR'16 dataset	Bi-LSTM	99%							

Shahid Baksh (2023)	IoT Network Security through DL-Powered IDS	CIC-IoT22 dataset	FFNN	99.93%							
			LSTM	99.85%							
			RNN	96.42%							
Hidasi et al. (2015)	Personalized Session-Based Recommendations with RNNs	RSC15 RecSys15	GRU	46.3% 69.2%							
Fanzhi Meng (2021)	GRU and Multi-autoencoder based Insider Threat Detection for Cyber Security	CERT v6.2, CERT v4.2	GRU		98.7% 99.1%						
Mohammed Nasser (2019)	Insider Threat Identification using GRU	CERT v4.2	GRU	92.0%							0.29
Alaca et al. (2023)	Graph-Based Log Anomaly Detection with GLSTM	HDFS, BGL, IMDB	GLSTM	97.01%	99.9%	99.8%			0.999	0.998	
Konstantinos Demertzis (2023)	CM-DANA for Anomaly Detection		CM-DANA	92%	88%				0.900		
Singh, M (2023)	User Behaviour based Insider Threat Detection using a Hybrid Learning Approach	CERT v4.2	ESN	98.7%					0.997	0.987	
Kim (2022)	Threat classification model for security information event management focusing on model efficiency		CNN-static (1D)	94.07%						0.951	

However, there still remains a gap in the literature concerning the exploration of hybrid models, such as the proposed (Attention-based Long Short, Term Memory) AB_LSTM, which combine the strengths of both LSTM networks and attention mechanisms. These models offer a promising direction for future research, particularly in their application to real-time cybersecurity monitoring.

Several studies showcased remarkable efficacy in their proposed optimizations. For instance, (Kiran Kumar, 2023) study on optimizing LSTM and Bi-LSTM models for crop yield prediction achieved notable percentages in accuracy, with Bi-LSTM models recording 94%, 72%, and 71% for wheat, groundnut, and barley predictions, respectively. Similarly, Zhanquan Wang's (2023) research in cybersecurity, employing Bi-LSTM models for malicious cyber-attack detection, achieved an impressive accuracy rate of 99%. These results underscore the potential of deep learning architectures, specifically Bi-LSTMs, in achieving high accuracy rates across different fields and datasets.

Also, comparing the performance metrics of optimized models with baseline models provides a comprehensive understanding of advancements. Shahid Baksh's 2023 research in IoT Network Security revealed that DL-powered IDS models, including FFNN, LSTM, and RandNN, surpassed baseline models, with FFNN achieving an accuracy rate of 99.93%, highlighting the superiority of these deep learning architectures in enhancing IoT security frameworks.

Moreover, a discernible trend across multiple studies is the dominance of LSTM and GRU variants in cybersecurity applications. For example, (Al-Mhiqani, 2020) study on Insider Threat Identification and Serrano's 2023 research on ESN for User Behavior Analytics in Cyber Threat Detection both utilized GRU architectures, indicating a prevailing preference for gated recurrent units in addressing cybersecurity challenges. Furthermore, the emergence of attention mechanisms, as evidenced by Vaswani et al.'s 2017 study on the "Attention Is All You Need" Transformer architecture, suggests a paradigm shift toward models capable of capturing intricate dependencies in sequential data. However, certain anomalies and disparities exist, such as the variance in datasets and evaluation metrics across studies, making direct comparisons challenging. For instance, while (Demertzis, 2023) study on CM-DANA for Anomaly Detection achieved an F1 score of 90%, the absence of such metrics in several

In addition, studies across various domains consistently showcase the prowess of deep architectures like LSTM and GRU, particularly in cybersecurity and behavior analytics tasks(Farhad, et al 2023, Ahmed, 2023). However, the observed inconsistencies in performance metrics across different datasets and scenarios underscore the latent influence of vanishing gradients. It is imperative that researchers should prioritize standardized evaluation metrics and benchmarks to ascertain the true impact of vanishing gradients. This would facilitate comparative analyses, foster transparency, and guide model selection tailored to specific tasks and datasets as also suggested by (Talaei Khoei T, 2023).

Certain studies have elucidated the efficacy of specific optimization techniques in mitigating the vanishing gradient challenge. Techniques such as weight sharing, batch normalization, and advanced weight initialization methods have demonstrated promising outcomes. Therefore, embracing a multi-pronged approach to optimization by integrating advanced techniques like bidirectional NNs and attention mechanisms can foster stable training dynamics, accelerate convergence, and enhance model robustness. Continuous experimentation and benchmarking are crucial to identify optimal combinations tailored to specific applications and architectures.

Notably, the depth and intricacy of neural network architectures significantly influence the manifestation and severity of vanishing gradients. Deeper architectures potentially capture intricate patterns but are inherently susceptible to this challenge. We therefore implorer researchers and practitioners to advocate for the development of hybrid architectures that judiciously balance depth and complexity. This approach can harness the representational prowess of deep networks while mitigating the inherent challenges associated with extensive architectures.

Overcoming Vanishing Gradient

The vanishing gradient problem remains a significant challenge in the application of traditional RNNs, particularly in deep networks. However, the literature highlights several successful approaches to mitigating this issue. The introduction of Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) has been pivotal in this regard. These architectures, by design, are more resistant to the vanishing gradient problem due to their gating mechanisms, which allow for better preservation and propagation of gradients during backpropagation. Furthermore, the use of advanced optimization techniques, such as the Adam optimizer and gradient norm regularization, has further enhanced the ability of these networks to learn long-term dependencies, which is crucial for effective cybersecurity applications.

Our study shows that incorporation of contextual features within datasets and the development of context-aware models emerge as pivotal strategies to address vanishing gradient challenges (Hu, 2018; Hu Z, 2021; Roodschild M, 2020). Contextual features provide nuanced insights, facilitating models to discern intricate patterns and relationships within data. Furthermore, the evolution towards dynamic models, capable of adapting to shifting contextual cues, amplifies model efficacy and robustness (Sivakrishna. K. R., 2023). Therefore, researchers should emphasize the integration of contextual features within datasets, fostering richer, more informative data representations. Such features empower models with enhanced discernment capabilities, enabling nuanced pattern recognition and prediction. Also, championing the development of context-aware and dynamic models is imperative. These models, equipped with adaptive mechanisms, harness contextual insights and dynamically recalibrate their predictions and inferences in response to evolving data landscapes. This agility augments model adaptability, accuracy, and resilience, especially in dynamic environments characterized by shifting patterns and relationships

The integration of LSTM networks with attention mechanisms has shown to be particularly effective in enhancing the contextual awareness of UEBA systems. Attention mechanisms allow the model to focus on relevant parts of the input sequence, thereby improving the interpretability and accuracy of threat detection. We therefore, propose leveraging an Attention-Based LSTM (AB-LSTM) model that incorporates context awareness features in datasets. AB-LSTM combines the strengths of LSTM for sequential modeling and attention mechanisms for focusing on relevant actions. By integrating context-aware features (such as user roles, access patterns, and network activity), the model can adapt to varying scenarios and identify subtle anomalies indicative of insider threats. The attention mechanism dynamically weighs context-aware features, allowing the model to prioritize critical information. This approach ensures fine-grained detection while considering the broader context in which user behaviors occur.

Given the standard Standard LSTM Notation as follows;

Input gate $\boldsymbol{\Gamma_i = \sigma(W_i \cdot [a_{t-1}, x_t] + b_i)}$, the forget gate $\boldsymbol{\Gamma_f = \sigma(W_f \cdot [a_{t-1}, x_t] + b_f)}$, the output gate $\boldsymbol{\Gamma_o = \sigma(W_o \cdot [a_{t-1}, x_t] + b_o)}$, the cell state update $\boldsymbol{\tilde{c}_t = \tanh (W_c \cdot [a_{t-1}, x_t] + b_c)}$, the final cell state $\boldsymbol{c_t = \Gamma_f \cdot c_{t-1} + \Gamma_i \cdot \tilde{c}_t}$ and the hidden state $\boldsymbol{a_t = \Gamma_o \cdot \tanh (c_t)}$.

We propose to stack LSTM with Attention mechanisms where the LSTM layer processes the input sequence (X = {x$_1$, x$_2$, ⨀ x$_T$}) and generates hidden states (H = {h_t, h_2, ⨀, h$_T$}).

The LSTM equations include:

$$\begin{aligned}
f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\
i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\
o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\
C_t &= \tanh (W_C \cdot [h_{t-1}, x_t] + b_C) \\
C_t &= f_t \cdot C_{t-1} + i_t \cdot C_t \\
h_t &= o_t \cdot \tanh (C_t)
\end{aligned}$$

The attention mechanism computes a context vector (C_t) that focuses on relevant parts of the input sequence. The attention score (e_t) is calculated as:

$$e_t = \tanh (W_e \cdot h_t + b_e).$$

The attention weights (α_t) are obtained using a softmax function:

$$\alpha_t = \frac{\exp (e_t)}{\sum_{k=1}^{T} \exp (e_k)}$$

Also, the context vector (c_t) is a weighted sum of the hidden states:

$$c_t = \sum_{k=1}^{T} \alpha_k \cdot h_k$$

And the final output (y_t) is generated by combining the context vector (c_t) and the hidden state (h_t):

$$y_t = \sigma(W_y \cdot [c_t, h_t] + b_y)$$

This model leverages the LSTM's ability to capture sequential dependencies and the attention mechanism's capability to focus on critical information, enhancing the detection of insider threats Therefore, the proposed AB_LSTM (Attention-Based LSTM) architecture builds on this concept by incorporating both temporal and contextual elements into the analysis, making it well-suited for the dynamic nature of cybersecurity threats. The AB_LSTM model demonstrates superior performance in capturing complex patterns in user behavior, indicating its potential for future research and application in advanced UEBA systems.

CONCLUSION AND FUTURE WORK

This systematic review underscores the importance of addressing the vanishing gradient problem in RNNs to enhance their application in cybersecurity, particularly in UEBA systems. The review has highlighted the effectiveness of advanced RNN architectures, such as LSTMs and GRUs, in mitigating this issue and improving the accuracy and reliability of threat detection. The integration of attention mechanisms into these architectures further enhances their contextual awareness, making them well-suited for dynamic and complex cybersecurity environments.

It therefore becomes evident that User and Entity Behavior Analytics (UEBA) stands at a critical juncture, demanding concerted research efforts to address its multifaceted challenges. A salient concern is the absence of rigorous comparative evaluations among dominant deep learning architectures, potentially steering organizations towards suboptimal UEBA solutions. This lacuna, coupled with the interpretability issues intrinsic to RNN-based models, underscores the urgency for enhanced transparency in AI-driven cybersecurity decision-making processes. As cyber threats burgeon in sophistication and scale, the imperatives of scalability, computational efficiency, and adaptability loom large, necessitating innovative architectural adaptations tailored explicitly for UEBA contexts and especially in a rapidly changing threat landscape.

To propel future research, it's imperative to prioritize models that are both contextually aware and dynamically adaptive. Contextual awareness ensures that UEBA systems comprehend organizational nuances, evolving threat landscapes, and user behaviors in specific contexts, thereby reducing false positives and enhancing anomaly detection accuracy. Concurrently, the development of dynamic models those that continuously learn, evolve, and adjust based on real-time data and shifting threat vectors is essential. Future research should explore the development and implementation of hybrid models, such as the proposed AB_LSTM, which combines the strengths of LSTM networks and attention mechanisms. These models offer promising potential for improving the accuracy and efficiency of UEBA systems in real-time threat detection.

Additionally, further investigation into alternative optimization techniques and their applicability to different cybersecurity scenarios will be essential in advancing the field and addressing emerging challenges in the ever-evolving cybersecurity landscape. Proposed innovations span attention mechanisms to prioritize pertinent user activities, hybrid convolutional-RNN architectures to capture temporal intricacies, and hierarchical models to discern anomalies at varying granularities. Additionally, integrating adaptive techniques, such as reinforcement learning or online learning, can enable UEBA systems to recalibrate and refine their detection capabilities proactively. As research advances, fostering interdisciplinary collaborations and prioritizing dynamic, contextually aware models will be paramount to fortify digital ecosystems, address organizational vulnerabilities, and advance the efficacy of RNN-based UEBA solutions in an ever-evolving cybersecurity landscape.

Implications for Cybersecurity

The findings of this review have significant implications for the development of more robust and effective cybersecurity measures. The optimized RNN models discussed, particularly those incorporating attention mechanisms, provide a powerful tool for enhancing UEBA systems' ability to detect and respond to insider threats. By improving the contextual awareness and long-term dependency modeling of these systems, organizations can achieve a higher level of security and threat detection accuracy. The proposed AB_LSTM model, in particular, holds potential for advancing the state-of-the-art in cybersecurity analytics by offering a more nuanced and contextually aware approach to threat detection.

Limitations

While the review provides valuable insights into RNN optimization methods, it is important to acknowledge certain limitations. The scope of the review was limited to published research available in selected databases, which may not fully represent the latest advancements in the field. Additionally, the focus on the vanishing gradient problem, while crucial, may have overlooked other significant challenges in RNN-based cybersecurity applications, such as computational efficiency and scalability. Future reviews could benefit from a broader scope and inclusion of a wider range of optimization challenges and solutions.

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to my supervisors, Dr. Sadiq O. Bashir, Dr. Bashir Olaniyi Sadiq, Ph.D., Dahiru Buhari, Ph.D., and Maninti Venkateswarlu, and my collegue Mr. Atuhe Aaron, for their invaluable guidance, support, and encouragement in this endeavor. Their expertise and constructive feedback have been instrumental in shaping my work and pushing me to achieve my best. I would also like to extend my deepest appreciation to my beloved wife, Nuwasasira Sharon. Her unwavering support and encouragement during the challenging moments, late nights, and moments of doubt have been a source of strength for me. Her understanding and patience have made this journey not only possible but also rewarding. I am profoundly grateful for her love and belief in my abilities, which have motivated me to persevere.

REFERENCES

1. Anna Fleck. (2024, Feb 22). Cybercrime Expected To Skyrocket in Coming Years. Retrieved from statista: https://www.statista.com/chart/28878/expected-cost-of-cybercrime-until-2027/
2. Jurgens, J. P. (2024). Global Cyber threat outlook. Switzerland: World Economic Forum.
3. Mathew, A. A. (2021). Advanced Machine Learning Technologies and Applications. Springer, Singapore.
4. Mathew, A. A. (2021). Advanced Machine Learning Technologies and Applications. Advances in Intelligent Systems and Computing. Singapore: Springer.
5. Fe A. Ajit, K. A., & Samanta. (2020). A review of convolutional neural networks. international conference on emerging trends in information technology and engineering (ic-ETITE), 2020: IEEE,, pp. 1-5.
6. Sen, A. T. (2021). A hybrid CNN-LSTM model for pre-miRNA classification. Scientific reports, 11(1).
7. Ahmed, S. A. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. rtif Intell Rev 56, 13521–13617. .
8. Al-Mhiqani, M. N. (2020). A Review of Insider Threat Detection: Classification, Machine Learning Techniques, Datasets, Open Challenges, and Recommendations. Applied Sciences 10, no. 15: 5208., 10(15).
9. Alaca, Y. Y. (2023). Anomaly Detection in Cyber Security with Graph-Based LSTM in Log Analysis. Chaos Theory and Applications.
10. Alamyar AM, L. w.-s. (2023). Detecting malicious attacks using Cyber-security models using Deep learning approach. Research Square.
11. Alzubaidi, L. Z. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Big Data, 8(53).
12. Bakhsh, S. A. (2023). Enhancing IoT network security through deep learning-powered Intrusion Detection System. Internet Things, 24.
13. Cremer, F. S. (2022). Cyber risk and cybersecurity: a systematic review of data availability. Geneva Pap Risk Insur Issues Pract 47, 698–736.
14. D. K. Rathore and P. K. Mannepalli. (2021). A Review of Machine Learning Techniques and Applications for Health Care. International Conference on Advances in Technology, Management & Education (ICATME), pp. 4-8.
15. D. Yu, H. W. (2021). "Mixed pooling for convolutional neural networks," in Rough Sets and Knowledge Technology:. 9th International Conference, RSKT (pp. pp. 364-375). Shanghai, China: Springer.
16. Das, S. T. (2023). Recurrent Neural Networks (RNNs): Architectures, Training Tricks, and Introduction to Influential Research. In: Colliot, O. (eds) Machine Learning for Brain Disorders. Neuromethods. 197. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-3195-9_4.
17. Demertzis, K. R. (2023). A Cross-Modal Dynamic Attention Neural Architecture to Detect Anomalies in Data Streams from Smart Communication Environments. Applied Sciences,, 13(17), 9648.

18. F. Meng, P. L. (2021). GRU and Multi-autoencoder based Insider Threat Detection for Cyber Security. Sixth International Conference on Data Science in Cyberspace (DSC) (pp. pp. 203-210). Shenzhen, China,: IEEE.
19. Brooks, C. (2023). Cybersecurity Trends & Statistics For 2023; What You Need To Know. Forbes,.
20. Goshisht, M. (2024). Machine Learning and Deep Learning in Synthetic Biology: Key Architectures, Applications, and Challenges. ACS Omega. 2024 Feb 19;9(9):9921-9945.
21. J, H. (2021). Cybercrime: Victimization, Perpetration, and Techniques. Am J Crim Just 46:837–842.
22. Hu Z, Z. J. (2021). Handling Vanishing Gradient Problem Using Artificial Derivative. IEEE Access 9:22371–22377.
23. Hu, Y. H. (2018). Overcoming the vanishing gradient problem in plain recurrent networks. arXiv preprint arXiv:1801.06105.
24. Sarker, I. H. (2021). Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Computer Science, , 2(6), p. 420, .
25. J. Gu et al. (2018). Recent advances in convolutional neural networks,. Pattern recognition, 77, 354-377.
26. X. Zhang, K. H. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 37(9), 1904-1916,.
27. Kim, J. &. (2022). Threat classification model for security information event management focusing on model efficiency. Computers & Security, 120, 102789.
28. Kingma, D. P. (2014). Adam: A Method for Stochastic Optimization. ArXiv. /abs/1412.6980 .
29. Kiran Kumar, V. R. (2023). Optimizing LSTM and Bi-LSTM models for crop yield prediction and comparison of their performance with traditional machine learning techniques. Appl Intell 53, 28291–28309.
30. L. Chen, S. L. (2021). Review of image classification algorithms based on convolutional neural networks,. Remote Sensing, 13(22), 4712.
31. Li W, M. W.-F. (2017). Enhancing collaborative intrusion detection networks against insider attacks using supervised intrusion sensitivity-based trust management model. Journal of Network and Computer Applications 77:135–145.
32. Lu W, L. J. (2021). A CNN-BiLSTM-AM method for stock price prediction. Neural Comput & Applic 33:4741–4753.
33. M. A. Wani, F. A. (2020). Advances in deep learning. Springer.
34. M. Shashanka, M. -Y. (2016). User and entity behavior analytics for enterprise security. IEEE International Conference on Big Data (Big Data) (pp. 1867-1874). Washington, DC, USA,: doi: 10.1109/BigData.2016.7840805. .
35. Mahmoud Abbasi, A. S. (2021). Deep Learning for Network Traffic Monitoring and Analysis (NTMA): A Survey,. Computer Communications ISSN 0140-3664, 170, 19-41.
36. Markus Riek, R. B. (2018). The costs of consumer-facing cybercrime: an empirical exploration of measurement issues and estimates. Journal of Cybersecurity, Volume 4(Issue 1).
37. Miftahutdinov, Z. &. (2019). Deep Neural Models for Medical Concept Normalization in User-Generated Texts. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop (pp. 393–399). Florence, Italy: Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-2055 .
38. N. B. Gaikwad, V. T. (2019). Efficient FPGA implementation of multilayer perceptron for real-time human activity classification. IEEE Access,. 7. IEEE Access, vol. 7, pp. 26696-26706.
39. Niu, Z. Z.-N. (2023). Recurrent attention unit: a new gated recurrent unit for long-term memory of important parts in sequential data. Neurocomputing, 517, 1-9. https://doi.org/10.1016/j.neucom.2022.10.050 .
40. Olaniyan R, R. S. (2023). Application of User and Entity Behavioral Analytics (UEBA) in the Detection of Cyber Threats and Vulnerabilities Management. Computational Intelligence for Engineering and Management Applications.
41. P. P. Shinde and S. Shah. (2018). A review of machine learning and deep learning applications. Fourth international conference on computing communication control and automation (ICCUBEA) (pp. 1-6.). IEEE.
42. Pascanu, R. M. (2013). On the difficulty of training Recurrent Neural Networks. arXiv:1211.5063v2

[cs.LG] .

43. Pascanu, R. M. (2012). Understanding the exploding gradient problem.
44. Noh, S.-H. (2021). nalysis of Gradient Vanishing of RNNs and Performance Comparison. Information 12:442., (p. https://doi.org/10.3390/info12110442).
45. Ribeiro, A. H. (2019). Beyond exploding and vanishing gradients: Analysing RNN training using attractors and smoothness.
46. Riek M, B. R. (2018). The costs of consumer-facing cybercrime: an empirical exploration of measurement issues and estimates. Journal of Cybersecurity.
47. Mathenge, R. (2023). Insider Threats in 2024: 30 Eye-Opening Statistics.
48. Jitpattanakul, S. M. (2021). Deep convolutional neural network with rnns for complex activity recognition using wrist-worn wearable sensor data. 10(14), p. 1685.
49. S. Squartini, A. H. (2003). Preprocessing based solution for the vanishing gradient problem in recurrent neural networks. Proceedings of the 2003 International Symposium on Circuits and Systems 2003. ISCAS '03. Bangkok, Thailand,: doi: 10.1109/ISCAS.2003.1206412.
50. Shashanka M, S. M.-Y. (2016). User and entity behavior analytics for enterprise security. IEEE International Conference on Big Data (Big Data). Washington DC,USA: IEEE.
51. Shiri, F. T. (2023). A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU. ArXiv abs/2305.17473.
52. Shrestha A, M. A. (2023). Review of Deep Learning Algorithms and Architectures. IEEE Access 7:53040–53065.
53. Sun, C. S. (2022). A Systematic Review of Echo State Networks From Design to Application. IEEE Transactions on Artificial Intelligence, 99.
54. Takudzwa Fadziso. (2020). Overcoming the Vanishing Gradient Problem during Learning Recurrent Neural Nets (RNN).
55. Talaei Khoei T, O. S. (2023). Deep learning: systematic review, models, challenges, and research directions. Neural Comput & Applic 35:23103–23124.
56. Van Houdt, G. M. (2020). A review on the long short-term memory model. Artif Intell Rev 53, 5929–5955.
57. Vaswani, A. N. (2017). Attention is All you Need. Neural Information Processing Systems .
58. W. Liu, Z. W. (2017). A survey of deep neural network architectures and their applications. Neurocomputing, 234, 11-26.
59. Wang, W. R. (2017). Deep convolutional neural networks for image classification: A comprehensive review. Neural computation, 29(9), 2352-2449.
60. Wenjuan Li, W. M.-F. (2017). Enhancing collaborative intrusion detection networks against insider attacks using supervised intrusion sensitivity-based trust management model. Journal of Network and Computer Applications. .
61. Paolo Dal Cin, J. J. (2023). World Economic Forum Global Cyber security Outlook.
62. Ying, X. (2019). An overview of overfitting and its solutions," in Journal of physics. 1168, 022022.
63. Yuhuang Hu, A. H.-C. (2019). Overcoming the vanishing gradient problem in plain recurrent networks. https://arxiv.org/abs/1801.06105, 20.
64. Z. Hu, J. Z. (2021). Handling Vanishing Gradient Problem Using Artificial Derivative. IEEE Access, 9, 22371-22377.
65. Zhao, Z. X. (2021). A LSTM-Based Anomaly Detection Model for Log Analysis. J Sign Process Syst 93,, 745–751.
66. Zhaoyang Niu, G. Z.-N. (2023). Recurrent attention unit: A new gated recurrent unit for long-term memory of important parts in sequential data. Neurocomputing, 517, 1-9.
67. Salitin MA, Z. A. (2023). The role of User Entity Behavior Analytics to detect network attacks in real time. IEEE, 1–5.
68. Roodschild M, G. S. (2020). A new approach for the vanishing gradient problem on sigmoid activation. Prog Artif Intell 9:351–360.
69. Jurgens, P. D. (2023). World Economic Forum. Global Cybersecurity Outlook. World Economic Forum.
70. Agarap, A. F. (2021). Deep Learning using Rectified Linear Units (ReLU). .
71. Bin Sarhan, B., & Altwaijry, N. (2023). Insider Threat Detection Using Machine Learning Approach.

Appl. Sci., 13, 259., 13.

72. Wang, A. M. (2023, July). Detecting malicious attacks using Cyber-security models using Deep learning approach.
73. Li, H. C. (2022). network, The architecture of dynamic reservoir in the echo state.
74. Singh, M. M. (2023). User Behaviour based Insider Threat Detection using a Hybrid Learning Approach. . Ambient Intell Human Comput , 14, 4573–4593.
75. Rafiq, G. R. (2023). A comprehensive survey of deep learning approaches. Artif Intell Rev 56, 13293–13372.
76. Andrea Galassi, M. L. (2021, October). Attention in Natural Language Processing. IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, 33(10).
77. Sivakrishna., K. R. (2023). An efficient pattern-based approach for insider threat classification using the image-based feature representation. Journal of Information Security and Applications.
78. Sivakrishna., K. R. (2023). An efficient pattern-based approach for insider threat classification using the image-based feature representation. Journal of Information Security and Applications