Application of a Recurrent Neural Network Model to Prevent Phishing Attacks: A Systematic Review, Challenges and Future Work

Carlos Oropeza [1], Alfredo Daza [1]

[1] *Faculty of Engineering and Architecture, School of Systems Engineering, Universidad César Vallejo, Lima, Peru*

Abstract **– Phishing is the most popular form of attacks in cyberspace, accounting for 1,270,883 of attacks on organizations. The main objective of the study is to understand Recurrent Neural Networks (RNN) applications in a solution to prevent phishing attacks. A systematic review of the literature was carried out to identify aspects such as: existing research on the types of RNN, their performance against different datasets and algorithms that complement this solution. For this, 30 articles were analyzed. Among the results obtained, it stands out that the most used type of RNN deep learning algorithm was Long Short-Term Memory (LSTM), the most used dataset was PhishTank, while LSTM was the model that obtained the best accuracy with a range of 92.12% to 99.86%, and Convolutional Neural Network (CNN) was the most used complementary algorithm to prevent phishing attacks. This research provides scientific evidence on how Deep Learning techniques can improve the detection of phishing attacks, contributing to the field of cybersecurity, in the prevention, detection and management of these attacks. In addition, it provides information to make more accurate decisions in the protection of computer systems, identifying gaps in the literature related to the prevention of phishing attacks.**

DOI: 10.18421/TEM144-07
https://doi.org/10.18421/TEM144-07

Corresponding author: Alfredo Daza,
Faculty of Engineering and Architecture, School of Systems Engineering, Universidad César Vallejo, Lima, Peru
Email: adaza@ucv.edu.pe

Received: 24 March 2025.
Revised: 26 April 2025.
Accepted: 05 May 2025.
Published: 27 November 2025.

Keywords **– Deep learning, recurrent neural network, phishing attacks, LSTM, cybersecurity.**

1. Introduction

Phishing is the most popular way to execute an attack in cyberspace that poses a risk to organizations, institutions, and individual users [1].

In the third quarter of 2022, the Anti-Phishing Working Group (APWG) published a report where a record number of phishing attacks was recorded, with a total of 1,270,883 attacks, of which 23.2% were directed at the financial sector, thus representing the most serious quarter recorded by the APWG [2]. On the other hand, Kaspersky recorded a 40% increase in phishing attacks in 2023, reaching a total of 709 million attacks, or 209 million more than in 2022 [3].

The main causes of phishing attacks are directly related to the human factor. These include a lack of knowledge about how these attacks work, a lack of awareness about the consequences of such a breach, and complacency stemming from a culture or work environment that overvalues cybersecurity policies, thus underestimating the real dangers of a phishing attack [4].

Phishing attacks often result in significant financial losses. According to a report published by Cloudflare in 2023, in the realm of corporate emails alone, phishing attacks have accounted for losses of more than $50 billion [5].

Another of the most common consequences of a phishing attack is the leakage of information. According to Verizon's Verizon 2024 Data Breach Investigations Report, 3,661 social engineering incidents were documented in 2023, where 3,032 incidents, or approximately 82.81%, involved a disclosure of information, of which 73% were related to phishing and email pretexting [6].

Due to the financial consequences that a phishing attack can cause due to information leaks, the healthcare sector becomes one of the most affected sectors.

According to a report by the International Business Machines Corporation (IBM) on the costs of information leaks in the year 2023, the healthcare sector has been the industry with the highest costs for 13 consecutive years, revealing an increase of 53.3% compared to 2020, with an average cost of 10.93 million dollars [7].

There are several strategies to avoid phishing attacks, including list-based methods, visual similarities, heuristic approach, the use of Machine Learning, Deep Learning models, and algorithms [8]. Deep Learning, in particular, is used to solve different problems observed in the health, industrial and information technology sectors, due to its ability to provide real-time results and its high accuracy in the detection of false positives, which increases user confidence [9].

The most commonly used Deep Learning models include CNNs (convolutional neural networks) and RNNs (recurrent neural networks). Within RNNs, there are variants, such as LSTM (long-term memory), GRU (Closed Recurrent Units), and BRNN (Bidirectional Recurrent Neural Networks) [10], [11], [12].

Several studies indicate that recurrent neural networks offer superior performance in various applications, however, it has not been the most preferred for phishing attack detection [13]. This is due to its limitations, such as the inability to address long-term dependencies, the problem of gradient fading, and especially, the high temporal complexity. These limitations are addressed through the use of RNN variants and/or hybrid models [14], [15].

This review is organized according to the various variants of RNNs previously Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bidirectional Recurrent Neural Network (BRNN). In addition, the use of hybrid approaches that combine RNN models with other techniques is addressed. Thus, the present study aims to understand RNN applications in a solution to prevent phishing attacks to identify aspects such as: existing investigations of RNN types, their performance against different datasets and algorithms that complement this solution. Based on that, it seeks to answer the following question: *"What is the state of the art of applying a recurrent neural network model to prevent phishing attacks?"*

The motivation for conducting this study is the importance of preventing phishing attacks and more specifically demonstrating how the application of a recurrent neural network affects these solutions.

This article will be divided into the following sections:

In the 2nd section the methodology of the systematic review will be discussed, where the methodology of the proposed study and the collection and analysis of the study articles will be explained, the 3rd section will focus on the results obtained after the analysis of the articles and the research questions will be answered, the 4th section will focus on the challenges and future work of the application of an RNN model for the prevention of phishing attacks, then the 5th section will be moved on to the discussion of the results obtained, and finally, the 6th section provides a series of conclusions from the research questions and from this systematic review.

2. Methodology

In this section, different points of the methodology proposed by [16] will be raised, which is focused on phishing detection based on deep learning, which will help to approach the systematic review at different points. The phases of this methodology are illustrated in Figure 1. First, the research questions and the strategy that was used for the search for articles will be presented, then the selection of articles, then it will be mentioned how the extraction was done and the respective analysis of the data obtained from the studies, then a quality evaluation of the articles will be carried out and finally, the inclusion and exclusion criteria taken into account in the research will be mentioned.

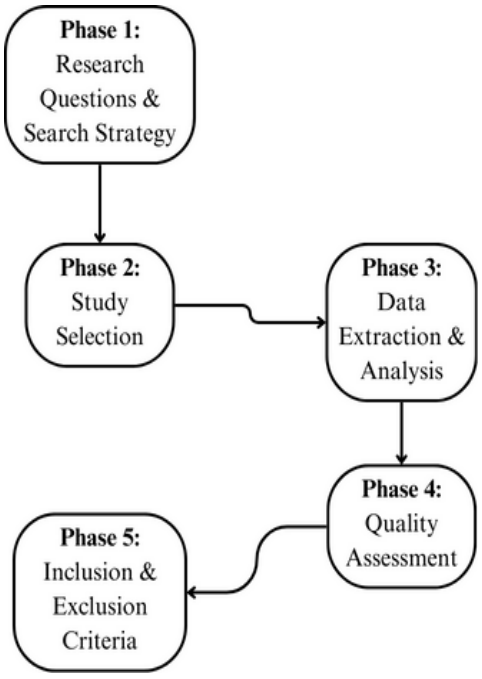

Figure 1. Phases of the methodology developed by [16].

2.1. Research Questions and Search Strategy

The research questions and search strategy phase consist of defining a clear question that will guide the study, as well as designing a strategy to find relevant literature, which involves selecting appropriate databases and selecting key terms related to the topic. [16].

In a systematic literature review (SLR), research questions are important to ensure that the review is comprehensive, based on solid evidence, and focused on the core objective of the study. The research questions asked for this study, along with their respective motivations, are as follows:

Table 1. Research questions and motivation

Research questions	**Motivation**
RQ1. What are the types of RNN deep learning algorithms that have been used in studies to prevent phishing attacks?	Identify the types of Deep Learning RNN algorithms used in the prevention of phishing attacks
RQ2. Which datasets were used to train and test the Deep Learning RNN models to prevent phishing attacks?	Discover the datasets to train and test deep learning RNN models to prevent phishing attacks
RQ3. What is the best accuracy value obtained in a Deep learning RNN model?	Reveal the accuracy of various Deep Learning RNN models.
RQ4. Which algorithms were used in conjunction with the Deep Learning RNN model to prevent phishing attacks?	Identify the complementary algorithms used in conjunction with the deep learning RNN model in the prevention of phishing attacks

As shown in Table 1, the first research question focused on the number of studies that propose a Deep learning model that uses recurrent neural networks or, failing that, recurrent neural network (RNN) variants, to prevent phishing attacks. The second research question discusses the datasets used to train and test models that use RNNs to prevent phishing attacks, to identify the source of datasets. The third question was posed to identify the RNN model, RNN variants and/or hybrid models where RNN is used, where the results obtained against the accuracy and precision of the models will be collected. And in the last question, the objective was to mention the algorithms that have been used with the RNN model to prevent phishing attacks.

To answer the research questions, a search strategy was carried out in various databases such as Scopus, ProQuest and WoS, using the following search string: “((“recurrent neural network" OR rnn) OR (lstm OR "Long short term memory”) OR (gru OR "Gated Recurrent Unit”) OR (brnn OR "Bidirectional recurrent neural networks”)) AND (“phishing attacks" OR "phishing detection" OR (phishing AND attacks AND detection) AND (“deep learning" OR dl))”.

After applying the search strategy using the chain in the databases mentioned above, a total of 1506 documents were obtained.

2.2. Selection of Studies

In the next phase, the studies were filtered, where only articles with less than 5 years of publication were included, in addition to being exclusively articles, when applying this filtering a total of 978 articles were obtained. Next, articles with open access were included, in addition to filtering only for Deep Learning and Phishing, thus obtaining a total of 129 articles. Giving importance to the quality of the present study, articles that are Q4 were excluded, thus leaving 113. The duplicate articles are then eliminated, resulting in 67 articles in total. Then, the titles and abstracts were reviewed to verify that they talk about the application of an RNN model to avoid phishing attacks, resulting in 34 articles. Finally, each article was reviewed in more detail in order to validate that the approach is relevant to the present research, resulting in 30 articles that answered the research questions.

2.3. Data Extraction and Analysis

The data was collected and in its subsequent analysis, the following relevant points were taken into account: what Deep Learning model was used (RNN, its variants or hybrid designs), datasets and their respective sources, algorithms with which was worked on with the solutions, what was innovated in their study, what were their limitations and what results were obtained.

2.4. Quality Assessment

In order to ensure the quality of the study, a thorough evaluation of the selected articles was carried out, based on five key criteria, which are detailed below:

QA1: Does the study focus directly on the analysis of phishing attacks using some type of RNN model, some variant or hybrid model?
QA2: Does the article meet the objective of this study?
QA3: Does the study answer any questions raised in the review?
QA4: Does the selected article specify all the points mentioned at the point of data extraction and analysis?
QA5: Does the article present a comparison of the proposed model with other relevant approaches?

Based on the aforementioned questions, if the article answers the question correctly, a maximum value of 1 is considered, in case it does not answer the question completely, a score of 0.5 is considered, and in case it does not answer the question in its entirety, it would represent a 0 to the score of the question. Table 2 presents the evaluation of the articles:

Table 2. Quality assessment by article

Ref.	QA1	QA2	QA3	QA4	QA5	Total
[17]	1	1	1	1	0.5	4.5
[18]	1	1	1	0.5	1	4.5
[19]	1	1	1	0.5	0.5	4
[20]	1	1	1	1	1	5
[21]	1	1	1	1	1	5
[22]	1	1	1	1	0.5	4.5
[23]	1	1	1	1	0	4
[24]	1	1	1	1	0.5	4.5
[25]	1	1	1	0.5	0.5	4
[26]	0.5	0.5	1	1	1	4
[27]	1	1	1	1	0.5	4.5
[28]	1	1	1	1	1	5
[29]	1	1	1	1	0.5	4.5
[30]	1	1	1	1	1	5
[31]	1	1	1	1	1	5
[32]	1	1	1	0.5	1	4.5
[33]	1	1	1	1	1	5
[34]	1	1	1	1	1	5
[9]	1	1	1	1	1	5
[35]	1	1	1	1	1	5
[36]	1	1	1	1	1	5
[37]	1	1	1	1	0.5	4.5
[38]	1	1	1	1	1	5
[39]	1	1	1	0.5	1	4.5
[40]	1	1	1	1	1	5
[41]	1	1	1	1	1	5
[42]	1	1	1	1	0.5	4.5
[43]	0.5	1	1	1	1	4.5
[44]	1	1	1	1	1	5
[45]	1	1	1	1	1	5

2.5. Inclusion and Exclusion Criteria

In order to ensure the relevance and quality of the selected studies, specific inclusion and exclusion criteria were established. These criteria help guide the selection process and guarantee that only the most pertinent literature related to the research topic is considered.

2.5.1. Inclusion Criteria

The inclusion criteria for this study required that the article be related to recurrent neural networks (RNN), their variants, or hybrid models incorporating RNN. Additionally, the article had to be published in English and obtained from reputable academic databases such as ProQuest, Scopus, or Web of Science. Finally, the selected studies needed to propose a solution addressing phishing attacks.

2.5.2. Exclusion Criteria

Articles were excluded if ranked below Q4, identified as duplicates, or categorized as theses, incomplete works, or books. Studies that were not aligned with the research questions or that did not fully adhere to the data extraction methodology specifically, lacking key elements required for the systematic review were also omitted. Furthermore, articles that did not present a results analysis were not considered for inclusion.

3. Results

According to the criteria and procedures carried out in the methodology point, in this section the results obtained were analyzed, where initially the distribution of the articles according to their years of publication was described, then each article was analyzed mentioning the journal and the quartile to which it corresponds, in addition to considering some innovations and limitations, then each research question was answered.

3.1. Distribution of Articles

A series of studies carried out mostly between 2023 and 2024 have been compiled, as shown in Figure 2, representing 66.67% of the total articles selected in this review. Analyzing the results, it is shown that there is a fairly recent interest in the use of RNN models focused on the prevention of phishing attacks.

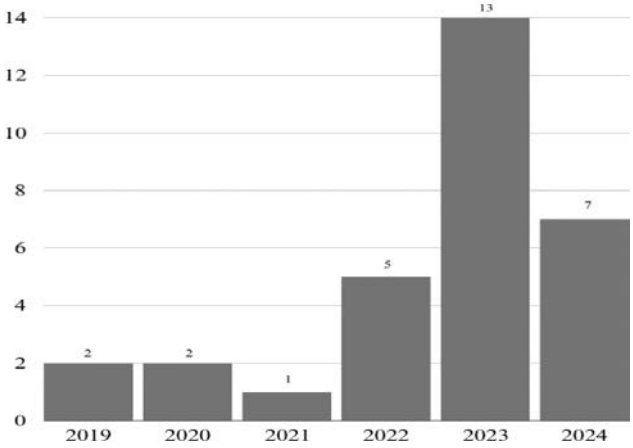

Figure 2. Number of articles published per year.

3.2. Analysis of Results

In contrast to the 30 articles considered in this research, the points mentioned in the data extraction and analysis section are analyzed, as presented in Table 3.

Table 3. Analysis of the articles (Innovation, limitations, journals and quartiles)

Ref.	Innovation	Limitations	Journal	Quartile
[17]	Combining BLSTM Model with RF	The model takes a long time to train large amounts of data and very small sample size	Complexity	Q3
[18]	Dataset balanced with a probabilistic GAN model type, in addition to introducing a binary adaptation of the AVOA.	CNN-LSTM model does not optimize parameters and does not specify size of Datasets	International Journal of Information Security	Q1
[19]	Use of Kalman filters in conjunction with a system of embedding words to LSTM and CNN models	Very small sample size and does not specify Dataset used	Applied Bionics and Biomechanics	Q3
[20]	Combining Datasets and applying LSTM with different approaches (bagging and stacking)	Little variety in the sample and does not apply the combined use of several Deep Learning methods	International Journal of Computer Networks and Communications	Q3
[21]	It uses the combination of the CNN and Bi-LSTM models with the use of IG (Information Gain) to classify characteristics	Very small sample size	International Journal of Electrical and Computer Engineering	Q2
[22]	Hybrid model working with CNN models, BiLSTM together with GHA (Gated Highway Attention)	Very small sample size, long training duration, incorrectly classifies abbreviated URLs.	e-Prime - Advances in Electrical Engineering, Electronics and Energy	Q3
[23]	Hybrid model using GRU, multiple layers of CNN and LSTM with large sample size	Low accuracy in sample lexical features	Baltic Journal of Modern Computing	Q3
[24]	Framework with hybrid model implementation and an algorithm for detecting multidimensional characteristics (statistics, code and text of the pages)	Does not apply Deep Learning to extract code and text from web pages	IEEE Access	Q2
[25]	Adaptable for numerical or text-based datasets. Can be used in CMD	Doesn't use many NLP techniques	Applied System Innovation	Q3
[26]	They apply a flexible hybrid model for multiple attacks including phishing	Small sample size and not 100% focused on phishing attacks	Sensors	Q2
[27]	Explores RNN model variants, demonstrates detailed results	It does not focus the study on a single model and the sample size is small	Future Internet	Q2
[28]	Filtered sample to maximize model accuracy	The model doesn't rank if a URL is real, so it affects detection accuracy	Security and Communication Networks	Q3
[29]	Large sample size with quality filtering, no reliance on third-party services	For future work, the complexity of the model is sought to be calculated for comparison	IEEE Access	Q2
[30]	Implement a hybrid model for detecting potential phishing attacks by voice recordings	Fairly small sample size, no results analysis in production	Mathematics	Q1
[31]	CNN model-based architecture working with LSTM, and in conjunction with a combination of CNN and LSTM.	Very small sample size	Electronics (Switzerland)	Q2
[32]	-	Does not specify sample size, errors in detecting non-English properties	Symmetry	Q2
[33]	Development of a phishing detection model focused on blockchain systems using a hybrid model	Very small sample size	Telecom	Q2
[34]	Application of Forward-Backwards LSTM model, making the solution robust and has a large sample size	Uses third-party services for page content detection	Computer Systems Science and Engineering	Q3
[9]	Using an RNN-GRU model, deploying the model to an extension in the Chrome web browser	Feature extraction with third-party services, web extension development is quite complex	IEEE Access	Q2
[35]	Robust framework using different models to assess the effects of the depth of a CNN model on phishing detection	Small sample size	Sensors	Q2
[36]	Accuracy: 95.14% (RF), Precision: 96.01% (NB), Accuracy: 95.13% (CNN), Accuracy: 95.14% (LSTM)	Detailed comparison of results with DL and ML models	Sensors	Q2
[37]	Using NLP with RNN variants for model development	Small sample size	Applied Sciences (Switzerland)	Q2

[38]	Application of a model using GRU and LSTM in multidimensional phishing detection	Small sample size, low result in Recall	TELKOMNIKA Telecommunication Computing Electronics and Control	Q2
[39]	Applying Machine Learning and Deep Learning together to develop a cloud solution	No sample size indicated	Complex & Intelligent Systems	Q2
[40]	Large, comparative sample size of all RNN variants, code, and Open Source Datasets	hardware limitations	IEEE Access	Q2
[41]	Apply strategy to target a lightweight model using GRU, ResNet, and ResNeXt. Implement an optimization approach	Small sample size	IEEE Access	Q2
[42]	Application of hybrid features considering text, images and fragments, making use of LSTM and CNN models	Detection takes a long time to detect each case (24 seconds).	Journal of Enterprise Information Management	Q1
[43]	Using DNN and LSTM models, using backpropagation to ensure model optimization	Small sample size	Neural Computing and Applications	Q2
[44]	Compare DNN, CNN, LSTM and GRU models in detail. It specifies the most common parameters and how the models studied act.	Small sample size	Applied Sciences (Switzerland)	Q2
[45]	The effectiveness of the different models for the detection of phishing in emails is evaluated	It must work with other security systems to provide more security	Electronics (Switzerland)	Q2

3.3. Answers to Research Questions

In this section, provided are answers to the research questions posed in this study by examining the findings from various sources.

3.3.1. Types of Deep Learning RNN Algorithms

Q1: What are the types of RNN deep learning algorithms that have been used in studies to prevent phishing attacks?

In the present research, a series of studies have been identified that apply RNN models to prevent phishing attacks, among them the following models are observed: Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (BiLSTM), Gated Recurrent Unit (GRU), Bidirectional Gated Recurrent Unit (BiGRU), and Bidirectional Recurrent Neural Network (BRNN). These models have been used as a single solution or, failing that, as a joint proposal forming a hybrid model.

For example, in the study by [35], a hybrid model called One-Dimensional Convolutional Neural Network for Phishing Detection (1D-CNNPD) is developed where the combination of the models was applied: CNN, LSTM and GRU together with their bidirectional variants. The use of a hybrid model generated a robust solution as indicated by researchers in the detection of phishing attacks through e-mails.

Similarly, in the article by [37], a solution that involves natural language processing and the use of the models was build: LSTM, BiLSTM, GRU and BiGRU, generating a model focused on preventing phishing attacks through URLs.

In the authors' research, Sahingoz, Buber and Kugu [40] present a solution for phishing detection using several models, which were: RNN, BRNN, CNN, ANN and an ATT, which were trained and tested with URLs. The experiments carried out by the research showed that the RNN-based models had good results regarding their accuracy, however, they were notably slower compared to the other models tested, in addition to mentioning that in more complex environments their accuracy was surpassed by CNN models.

The aforementioned research provides a general idea of how RNN models behave in different scenarios. Each specific reference is shown in Table 4, for consultation in greater detail.

Table 4. RNN models used in the articles

Model	**Ref.**	**Quantity (%)**
RNNs	[28], [40], [45]	3 (10.00%)
LSTM	[18], [19], [20], [23], [24], [25], [26], [27], [29], [31], [32], [33], [34], [35], [36], [37], [38], [39], [42], [43], [44], [45]	22 (73.30%)
BiLSTM	[17], [21], [22], [27], [30], [33], [34], [35], [37], [43]	10 (33.33%)
GRU	[9], [23], [27], [35], [37], [38], [41], [44]	8 (26.67%)
BiGRU	[35], [37]	2 (6.67%)
BRNN	[40]	1 (3.33%)

Based on the table presented above, the LSTM variant is used in almost all articles, having a total of 22 articles where this model is applied. This is due to the various features provided by LSTM, allowing to solve the fading and explosion of gradients existing in simple RNN models.

Gradient explosion and fading refers to a problem in the training stage of a simple RNN model, which is initially observed in the case of gradient explosion in the exponential growth of long sequences of data points compared to short sequences. And, on the other hand, with regard to gradient fading, it focuses on the exponential approach to 0 in long sequences of data points compared to shorter sequences [46].

3.3.2. Dataset for training and testing RNN models

Q2: Which datasets were used to train and test the Deep Learning RNN models to prevent phishing attacks?

The training and testing of the models obtained from the studies collected in this research are carried out through datasets from different sources, which depending on the type of data, item or approach would respond to their most common use or confidence within the studies collected and analyzed.

For example, in the article by [9], a hybrid RNN-GRU model was used in conjunction with ML algorithms such as RF (Random Forest) and LR (Logistic Regression) applying the Datasets: Phish storm, Phish Tank, ISCX-URL2016 and Kaggle, having a total of 429,125 legitimate URLs obtained from Phish storm (48,009 URLs), ISCX-URL2016 (35,378 URLs) and Kaggle (345,738 URLs), and 236,362 Phishing URLs obtained from Phish storm (47,902 URLs), Phish Tank (178,495 URLs) and ISCX-URL2016 (9,965 URLs) totaling 665,487 URLs. Obtaining 99.18% accuracy.

In addition, the research [34] proposes a solution making use of a CNN, an LSTM and its bidirectional variant, where the following Datasets are applied for training and testing: PhishTank and DomCop, having a total of 1,140,599 legitimate URLs and 1,167,201 phishing URLs, adding up to a total of 2,307,800 URLs. Obtaining a total of 98.5% accuracy.

Table 5. Datasets and article web pages

Dataset	Web page	Ref.	Quantity (%)
PhishTank	https://www.phishtank.com/	[9], [18], [20], [21], [23], [24], [28], [29], [32], [34], [38], [40], [42]	13 (43.33%)
Kaggle	https://www.kaggle.com/search	[9], [27], [41], [22]	4 (13.33%)
DomCop	https://www.domcop.com/	[21], [29], [34]	3 (10.00%)
SpamAssasin	https://spamassassin.apache.org/	[35], [45]	2 (6.67%)
UCI	https://archive.ics.uci.edu/	[44], [45]	2 (6.67%)
ISCX 2016	https://www.unb.ca/cic/datasets/url-2016.html	[9], [31]	2 (6.67%)
CommonCrawl	https://commoncrawl.org/overview	[40], [42], [32]	3 (10.00%)
Top 1 Million domain of the alexa	https://www.kaggle.com/datasets/cheedcheed/top1m	[28], [32]	2 (6.67%)
Phishing website dataset (UCI)	https://archive.ics.uci.edu/	[17]	1 (3.33%)
KorCCVi v2	https://github.com/selfcontrol7/Korean_Voice_Phishing_Detection/tree/main/Attention	[30]	1 (3.33%)
Ethereum transaction network	https://github.com/MyEtherWallet/ethereum-lists	[33]	1 (3.33%)
Phishload	https://www.medien.ifi.lmu.de/team/max.maurer/files/phishload/	[37]	1 (3.33%)
Otros Datasets	-	[25], [18], [32], [39]	4 (13.33%)
No especifica	-	[19]	1 (3.33%)

On the other hand, in the article by. [31] propose a hybrid model working with LSTM and CNN, the research used the ISCX-URL2016 Dataset containing 20,000 URLs, of which 10,200 URLs were legitimate and 9,800 URLs were phishing. Obtaining 99.2% accuracy.

Based on the above-mentioned articles, it can be denoted that the datasets are varied and taken from different sources together.

Table 5 shows in detail the datasets used by the research together with the links of each of the sources.

Based on the information presented in Table 4, 13 articles that make use of the PhishTank source are observed, followed by Kaggle having 4 articles in common that make use of this data source. The use of the PhishTank source in a large number of studies is due to its workflow and the passage of time it has generated that it will be standardized as the main source of data related to Phishing. In addition, this website allows the reporting and validation of Phishing URLs through the same community, thus generating a very reliable service.

3.3.3. Accuracy of RNN Models

Q3: What is the best accuracy value obtained in a Deep learning RNN model?

Each article that proposes a model as a solution against phishing attacks selected in this review, represents according to its metrics how accurate it is when making a prediction, this is influenced by variables such as sample size, sample quality, data preprocessing and support algorithms for the solutions.

For example, the research of [24], make use of a multidimensional feature algorithm with a CNN-LSTM model, which is trained and tested with 2,007,779 Uniform Resource Locators (URLs), of which 1,021,758 URLs are obtained from the PhishTank dataset and are phishing in their entirety, and on the other hand, the remaining 986,021 URLs are obtained from the dmoztools.net dataset and are legitimate in their entirety.

This article proposed a data validation and separation strategy for training and testing, where it considered two subsets of data, one where the URLs that could be accessed and, therefore, perform the extraction of content from the web made up of 44,835 (22,445 phishing and 22,390 legitimate), and the other subset of data includes the inaccessible URLs that would be used to train the effectiveness of the CNN-LSTM model made up of 1,965,944 URLs (999,313 phishing and 966,631 legitimate). The model proposed by the research obtained 98.99% accuracy, being the best result in relation to the sample size and its comparison with studies related to this review.

On the other hand, [19] proposes the use of two models, LSTM and a CNN, also making use of a sequential filtering called "Kalman" that allowed feeding the aforementioned neural learning architectures. This research used a Dataset with 11,430 URLs, of which half are phishing and the other half are legitimate. The results against the use of the two main models, i.e., LSTM and CNN, were 92.12% and 91.47% respectively. However, the use of the filtering proposed by the research considerably increased the result of both models, obtaining 99.86% for the LSTM model and 99.26% for the CNN model.

The study by [35], presented a hybrid model that made use of the models: CNN, LSTM, Bi-LSTM, GRU and its bidirectional variant. The Datasets used in the solution were Phishing Corpus and Spam Assassin Scheme, which together collected 13,362 emails, of which 9,212 were phishing and 4,150 were legitimate. The model got 99.68% accuracy.

Despite most results representing extremely high values with loss rates close to 0%, these results will always be influenced by the size of the Dataset used and the quality of the data in it. Table 6 shows the accuracy results for each RNN model.

Table 6. Accuracy and precision of the models used in the articles

Model	Accuracy	Precision	Ref.
BiLSTM	95.47%	95.60%	[17]
CNN -LSTM	99.37%, 98.87%, 98.79%	99.11%	[18]
LSTM-Kalman, LSTM	99.86%, 92.12%	99.90%, 92.14%	[19]
LSTM (bagging), LSTM (stacking)	99.62%, 99.65%	99.86%, 99.81%	[20]
CNN-BiLSTM	99.27%	98.95%	[21]
BiLSTM-GHA-CNN	99.78%	98.97%	[22]
GRU, LSTM	94.4%	97.16%	[23]
CNN-LSTM	98.61%	99.41%	[24]
LSTM	99%	99%	[25]
CNN-LSTM	98.74%, 98.29%	87.20%, 95.82%	[26]
LSTM, BiLSTM, GRU	96.9%, 99%, 97.5%	96.50%, 99%, 97%	[27]
RNNs	95.6%	97.33%	[28]
LSTM	97.58%	98%	[29]
CNN-BiLSTM	99.32%	99.32%	[30]
LSTM, LSTM-CNN	99.2%	99%	[31]
LSTM	96.04%	98%	[32]
LSTM, BiLSTM, CNN-LSTM	99.72%	99.86%	[33]
LSTM	98.5%	98.8%	[34]
RNN-GRU	99.18%	98.96%	[9]
LSTM, BiLSTM, GRU, BiGRU	99.68%	100%	[35]
LSTM	95.14%	87.28%	[36]
LSTM, BiLSTM, GRU, BiGRU	97.39%	-	[37]
LSTM, GRU	98.89%	99.17%	[38]
LSTM	98%	98%	[39]
RNN, BRNN	95.1%	-	[40]
GRU	98%	97.80%	[41]
LSTM	93.28%	93.30%	[42]
DNN-LSTM	98.79%, 99.21%	-	[43]
LSTM, GRU	97.20%, 96.70%	96.45%, 96.06%	[44]
RNN, LSTM, BERT-LSTM	98.58%, 98.89%, 99.61%	98.57%, 98.87%, 99.87%	[45]

The data shown above in Table 5 show the approximate accuracy results in reference to the different models of RNNs and their variants. It is remarkable the high percentages of accuracy obtained by solutions that make use of LSTM, GRU, in addition to their bidirectional variants. The bidirectional variant of LSTM is the one that obtains the best results in reference to the accuracy of the models that apply it.

3.3.4. Algorithms Used in Conjunction with the RNN Model

Q4: Which algorithms were used in conjunction with the Deep Learning RNN model to prevent phishing attacks?

The development of a solution against a phishing attack using an RNN model usually uses some algorithms that allow improving the results of these models, which can be focused on the preprocessing of the data, some feature selection layers, and even for the prediction layer of a model.

For example, the research of [43], makes use of natural language processing (NLP) in conjunction with some ML methods such as: k-Nearest Neighbors (kNN), support vector machine (SVM), in addition to some other tree-based methods, and DL algorithms such as DNN and a BiLSTM. Describes the use of ML algorithms to obtain higher accuracy results and improve dataset partitioning. Next, the solution, through the use of the Deep Learning algorithms used, generates a hybrid model that would be compared with the classification of the counterpart that is designed from the NLP model, concluding in a more accurate prediction. Theuse of the Datasets was made: Ebbu2017 and PhishTank that are used separately for comparison. Obtaining 98.79% against the Ebbu2017 Dataset and 99.21% in the PhishTank Dataset.

The use of varied algorithms that support different layers of a solution is necessary to obtain more accurate results when it comes to preventing phishing attacks. Table 7 shows the top 5 most commonly used algorithms in the development of an RNN model (variants or hybrid models).

Table 7. Top 5 algorithms used in the selected studies

Algorithm	Ref.	Quantity (%)
CNN	[19], [21], [22], [23], [24], [26], [28], [29], [30], [31], [33], [34], [35], [36], [40], [42], [44], [45]	18 (60.00%)
NLP	[23], [25], [37], [43], [45]	5 (16.67%)
RF	[17], [9], [36]	3 (10.00%)
SVM	[39], [43]	2 (6.67%)
GAN	[18], [29]	2 (6.67%)

The data observed in Table 6 notably demonstrate that convolutional neural networks (CNNs) are the most used with 18 studies that include this algorithm in conjunction with solutions that apply an RNN model or, failing that, some variant. CNN models are often used quite frequently due to their ability to analyze text-based data, such as a URL or text within an email. In the same way, this type of algorithms is chosen due to their high performance in modern solutions [45], [47].

Next, the second most used algorithm with 5 studies that include it, is natural language processing (NLP), which is usually chosen for its focus on understanding text in the way a person would, which is why when dealt with texts, such as URLs, the use of NLP allows the extraction of numerous characteristics for the model of a solution [48].

The third most used algorithm with 3 studies that include it within its solution is the Random Forest (RF), which is used due to its ability to combine several tree-type data structures, which separately generate results that are then combined and, in this way, generate a prediction, in this way, increasing the accuracy of the model in question [49].

Then, the fourth most used algorithm with 2 studies in common that involve it, is the support vector machine (SVM), which is used in certain solutions since it is a powerful classifier focused on texts, thus improving the model's ability to understand a set of data to what could be labeled as a URL or even text within an email [50].

Finally, the fifth algorithm has 2 studies in common; the generative adversarial networks. These are used within a phishing detection solution due to their ability to collect and learn data from texts such as a URL [51].

4. Challenges and Future Work

Through the analysis of the articles collected throughout this systematic review, some challenges are evident in the development of a solution against phishing attacks using an RNN model. Initially, it is observed that the quality of the data represents a challenge for research, because the obtainment, labeling and data variability is quite weak, due to the fact that the sample sizes are mostly small, in addition to samples without excessive variability and even data losses due to labeling errors. On the other hand, it is observed that the training and testing of the models focuses on characteristics of a single language, thus representing a challenge for the proposed models, because they would be incapable of being applied in environments that involve a language different from the one that was trained and tested.

Likewise, the training of models that use RNNs, a variant of RNNs or a hybrid model represent in some cases a challenge with respect to the high training times required by the models.

Finally, the dependence of the proposed models on working with third-party services to fully realize the solution is also considered a challenge.

The challenges presented above represent aspects of improvement in the models, which is why in this research some future work is proposed that would allow solving the challenges commented on. A proposal should initially focus on the quality of the data, focusing directly on explaining in detail the size of the sample, the quality of the sample, and also how it is being validated, considering in the same way that the dataset is large. To improve a proposal that has as its main objective to reduce phishing attacks, it must be developed and focused on the understanding of several languages.

On the other hand, although it is true that the RNN models throughout the present research have shown a fairly positive performance, on the side of training time it has not presented the same results, it is because of this, that future works should focus their solution on optimizing and thus improving the performance of the model so that the training time can be reduced considerably. Ultimately, the work proposed for future research should focus on custom development, without the use of third-party tools that are adapted to the solution, in addition to developing multidimensional models that are adaptable to different phishing attacks considering the points mentioned above.

5. Discussion

Thanks to the results obtained, presented and analyzed previously in the review article, different variables have been observed with respect to the research questions that were posed. In the first question, it is possible to observe a large number of studies that focused their solution on variants of the RNN architecture, more specifically, LSTM, GRU and their bidirectional variants. On the other hand, in the second research question, the use of Datasets such as PhishTank, and DomCop, which provide a large amount of data, in addition to making use of repositories hosted on Kaggle as samples for their research, is largely visualized.

In the third question posed in the research, it can be interpreted that the use of models such as LSTM, GRU and hybrid variants have high percentages of accuracy and precision. Finally, in the fourth question, the constant use of convolutional neural networks (CNNs), also the use of natural language processing (NLP) algorithms and finally, some ML algorithms such as: Random Forest (RF) and support vector machines (SVM) as complement algorithms in solutions against phishing attacks using RNN models.

6. Conclusions

This systematic review has been able to identify and answer the research questions, in addition to presenting the challenges and future approaches to solutions regarding the detection of phishing attacks using an RNN model, a variant of RNNs or, failing that, a hybrid model. It is important to mention the quite positive performance of the application of a recurrent neural network model in this type of problem presented throughout the research. It was identified that the use of variants of RNNs has been the focus of research to solve phishing attacks. On the other hand, the existence of small sample sizes was observed in most of the investigations, thus representing a challenge for future work.

The proposed models presented high results regarding prediction accuracy, where particularly a hybrid model that made use of CNN-LSTM presented the highest result in relation to the sample size and the different metrics of the model, obtaining 98.61% accuracy and 99.41% accuracy. On the other hand, regarding the algorithms that work together with an RNN model, it was mostly observed that it was common to use a solution where they included a CNN model, in addition to some ML-based layer support algorithms.

In summary, the development of this research raised different questions that answered the current state of the art of how the application of an RNN model prevents a phishing attack, detailing the challenges and future work that will allow new research to be proposed for the benefit of better models for the prevention of phishing attacks.

References:

[1]. Halgaš, L., Agrafiotis, I., & Nurse, J. R. (2019). Catching the phish: Detecting phishing attacks using recurrent neural networks (rnns). *International Workshop on Information Security Applications*, 219-233. Springer.

[2]. Anti-Phishing Working Group. (2022). *Phishing Activity Trends Report* (3rd Quarter). APWG. Retrieved from: https://docs.apwg.org/reports/apwg_trends_report_q3_2022.pdf [accessed: 02 February 2025].

[3]. Kulikova, T., et al. (2024). *El spam y el phishing en 2023.* Securelist. Retrieved from: https://securelist.lat/spam-phishing-report-2023/98496 [accessed: 03 February 2025].

[4]. Desolda, G., et al. (2021). Human factors in phishing attacks: a systematic literature review. *ACM Computing Surveys (CSUR)*, *54*(8), 1-35. Doi: 10.1145/3469886

[5]. Dzuba, E., & Cash, J. (2023). *Informe de Cloudflare sobre las amenazas de phishing en 2023. Cloudflare.* Retrieved from: https://blog.cloudflare.com/es-es/2023-phishing-report/ [accessed: 04 February 2025].

[6]. Hylender, D., et al. (2024). *2024 Data Breach Investigations Report*. Verizon business. Retrieved from: https://www.verizon.com/business/en-au/resources/reports/2024/dbir/2024-dbir-data-breach-investigations-report.pdf [accessed: 15 February 2025].

[7]. IBM. (2023). Informe sobre el costo de una filtración de datos 2023 Resumen Ejecutivo. Retrieved from: https://www.ibm.com/es-es/reports/data-breach [accessed: 02 March 2025].

[8]. Safi, A., & Singh, S. (2023). A systematic literature review on phishing website detection techniques. *Journal of King Saud University-Computer and Information Sciences*, *35*(2), 590-611. Doi: 10.1016/j.jksuci.2023.01.004

[9]. Tang, L., & Mahmoud, Q. H. (2021). A deep learning-based framework for phishing website detection. *IEEe Access*, *10*, 1509-1521. Doi: 10.1109/ACCESS.2021.3137636

[10]. Hamayel, M. J., & Owda, A. Y. (2021). A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms. *AI*, *2*(4), 477-496. Doi: 10.3390/ai2040030

[11]. Subramanian, B., et al. (2022). An integrated mediapipe-optimized GRU model for Indian sign language recognition. *Scientific Reports*, *12*(1), 11964. Doi: 10.1038/s41598-022-15998-7

[12]. Zhang, Y. (2023). Traffic flow prediction based on transformer. *Journal of Physics: Conference Series*, *2649*(1), 1-10. Doi: 10.1088/1742-6596/2649/1/012064

[13]. Catal, C., et al. (2022). Applications of deep learning for phishing detection: a systematic literature review. *Knowledge and Information Systems*, *64*(6), 1457-1500. Doi: 10.1007/s10115-022-01672-x

[14]. Dong, J., Chen, Y., & Guan, G. (2020). Cost index predictions for construction engineering based on LSTM neural networks. *Advances in Civil Engineering*, *2020*(1), 6518147. Doi: 10.1155/2020/6518147

[15]. Lavanya, P., Reddy, I. S., & Selvakumar, V. (2024). Long range radio technology implementation on internet of things to detect particulate matter at the community level and prediction using machine learning based approach. *Engineered Science*, *29*, 1119. Doi: 10.30919/es1119

[16]. Thakur, K., et al. (2023). A systematic review on deep-learning-based phishing email detection. *Electronics*, *12*(21), 4545. Doi: 10.3390/electronics12214545

[17]. Wang, S., et al. (2020). Deep learning-based efficient model development for phishing detection using random forest and BLSTM classifiers. *Complexity*, *2020*(1), 8694796. Doi: 10.1155/2020/8694796

[18]. Elberri, M. A., et al. (2024). A cyber defense system against phishing attacks with deep learning game theory and LSTM-CNN with African vulture optimization algorithm (AVOA). *International Journal of Information Security*, *23*(4), 2583-2606. Doi: 10.1007/s10207-024-00851-x

[19]. Tang, Y., & Wu, F. (2022). A Deep Learning Filter that Blocks Phishing Campaigns Using Intelligent English Text Recognition Methods. *Applied Bionics and Biomechanics*, *2022*(1), 5036026. Doi: 10.1155/2022/5036026

[20]. Banik, B., & Sarma, A. (2023). Phishing URL detection using LSTM based ensemble learning approaches. *International Journal of Computer Networks & Communications*, *15*(1), 17-33. Doi: 10.5121/ijcnc.2023.15102

[21]. Salah, H., & Zuhair, H. (2023). Deep learning in phishing mitigation: a uniform resource locator-based predictive model. *International Journal of Electrical and Computer Engineering*, *13*(3), 3227-3243. Doi: 10.11591/ijece.v13i3.pp3227-3243

[22]. Nanda, M., Saraswat, M., & Sharma, P. K. (2024). Enhancing Cybersecurity: A Review and Comparative Analysis of Convolutional Neural Network Approaches for Detecting URL-Based Phishing Attacks. *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, *8*(1), 100533. Doi: 10.1016/j.prime.2024.100533

[23]. Rasymas, T., & Dovydaitis, L. (2020). Detection of phishing URLs by using deep learning approach and multiple features combinations. *Baltic journal of modern computing.*, *8*(3), 471-483. Doi: 10.22364/bjmc.2020.8.3.06

[24]. Yang, P., Zhao, G., & Zeng, P. (2019). Phishing website detection based on multidimensional features driven by deep learning. *IEEE access*, *7*, 15196-15209. Doi: 10.1109/ACCESS.2019.2892066

[25]. Dewis, M., & Viana, T. (2022). Phish responder: A hybrid machine learning approach to detect phishing and spam emails. *Applied System Innovation*, *5*(4), 73. Doi: 10.3390/asi5040073

[26]. Djenna, A., et al. (2023). Unmasking cybercrime with artificial-intelligence-driven cybersecurity analytics. *Sensors*, *23*(14), 6302. Doi: 10.3390/s23146302

[27]. Roy, S. S., et al. (2022). Multimodel phishing URL detection using LSTM, bidirectional LSTM, and GRU models. *Future Internet*, *14*(11), 340. Doi: 10.3390/fi14110340

[28]. Wang, W., et al. (2019). PDRCNN: Precise phishing detection with recurrent convolutional neural networks. *Security and Communication Networks*, *2019*(1), 2595794. Doi: 10.1155/2019/2595794

[29]. Al-Ahmadi, S., Alotaibi, A., & Alsaleh, O. (2022). PDGAN: Phishing detection with generative adversarial networks. *IEEE Access*, *10*(1), 42459-42468. Doi: 10.1109/ACCESS.2022.3168235

[30]. Moussavou Boussougou, M. K., & Park, D. J. (2023). Attention-based 1D CNN-BiLSTM hybrid model enhanced with fasttext word embedding for Korean voice phishing detection. *Mathematics*, *11*(14), 3217. Doi: 10.3390/math11143217

[31]. Alshingiti, Z., et al. (2023). A deep learning-based phishing detection system using CNN, LSTM, and LSTM-CNN. *Electronics*, *12*(1), 232. Doi: 10.3390/electronics12010232

[32]. Aslam, S., et al. (2024). AntiPhishStack: LSTM-based stacked generalization model for optimized phishing URL detection. *Symmetry*, *16*(2), 248. Doi: 10.3390/sym16020248

[33]. Ogundokun, R. O., et al. (2023). Phishing detection in blockchain transaction networks using ensemble learning. *Telecom*, *4*(2), 279-297. Doi: 10.3390/telecom4020017

[34]. Tenis, A. A., & Santhosh, R. (2023). Modelling an Efficient URL Phishing Detection Approach Based on a Dense Network Model. *Computer Systems Science & Engineering*, *47*(2), 2625–2641. Doi: 10.32604/csse.2023.036626

[35]. Altwaijry, N., et al. (2024). Advancing Phishing Email Detection: A Comparative Study of Deep Learning Models. *Sensors*, *24*(7), 2077. Doi: 10.3390/s24072077

[36]. Nagy, N., et al. (2023). Phishing URLs detection using sequential and parallel ML techniques: comparative analysis. *Sensors*, *23*(7), 3467. Doi: 10.3390/s23073467

[37]. Benavides-Astudillo, E., et al. (2023). A phishing-attack-detection model using natural language processing and deep learning. *Applied Sciences*, *13*(9), 5275. Doi: 10.3390/app13095275

[38]. Subhashini, N., et al. (2024). Deep learning based phishing website detection. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *22*(1), 113-121. Doi: 10.12928/telkomnika.v22i1.25516

[39]. Butt, U. A., et al. (2023). Cloud-based email phishing attack using machine and deep learning algorithm. *Complex & Intelligent Systems*, *9*(3), 3043-3070. Doi: 10.1007/s40747-022-00760-3

[40]. Sahingoz, O. K., Buber, E., & Kugu, E. (2024). Dephides: Deep learning based phishing detection system. *IEEE Access*, *12(*1), 8052-8070. Doi: 10.1109/ACCESS.2024.3352629

[41]. Alsubaei, F. S., Almazroi, A. A., & Ayub, N. (2024). Enhancing phishing detection: A novel hybrid deep learning framework for cybercrime forensics. *IEEE Access*, *12*(1), 8373-8389. Doi: 10.1109/ACCESS.2024.3351946

[42]. Adebowale, M. A., Lwin, K. T., & Hossain, M. A. (2023). Intelligent phishing detection scheme using deep learning algorithms. *Journal of Enterprise Information Management*, *36*(3), 747-766. Doi: 10.1108/JEIM-01-2020-0036

[43]. Ozcan, A., et al. (2023). A hybrid DNN–LSTM model for detecting phishing URLs. *Neural Computing and Applications*, *35*(7), 4957–4973. Doi: 10.1007/s00521-021-06401-z

[44]. Do, N. Q., et al. (2021). Phishing webpage classification via deep learning-based algorithms: an empirical study. *Applied Sciences*, *11*(19), 9210. Doi: 10.3390/app11199210.

[45]. Atawneh, S., & Aljehani, H. (2023). Phishing email detection model using deep learning. *Electronics*, *12*(20), 4261. Doi: 10.3390/electronics12204261

[46]. Tsantekidis, A., Passalis, N., & Tefas, A. (2022). Recurrent neural networks. *Deep Learning for Robot Perception and Cognition*, 101-115.

[47]. Gupta, J., Pathak, S., & Kumar, G. (2022). Deep learning (CNN) and transfer learning: a review. *Journal of Physics: Conference Series*, *2273*(1), 012029. Doi: 10.1088/1742-6596/2273/1/012029

[48]. Elsadig, M., et al. (2022). Intelligent deep machine learning cyber phishing url detection based on bert features extraction. *Electronics*, *11*(22), 3647. Doi: 10.3390/electronics11223647

[49]. Sindhu, S., et al. (2020). Phishing detection using random forest, SVM and neural network with backpropagation. *International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, 391-394. Doi: 10.1109/ICSTCEE49637.2020.9277256

[50]. Orunsolu, A. A., Sodiya, A. S., & Akinwale, A. T. (2022). A predictive model for phishing detection. *Journal of King Saud University-Computer and Information Sciences*, *34*(2), 232-247. Doi: 10.1016/j.jksuci.2019.12.005

[51]. Kamran, S. A., Sengupta, S., & Tavakkoli, A. (2021). Semi-supervised conditional GAN for simultaneous generation and detection of phishing URLs: A game theoretic perspective. *arXiv preprint arXiv:2108.01852*. Doi: 10.48550/arXiv.2108.01852