electronics

MDPI

Article

A Deep Learning-Based Phishing Detection System Using CNN, LSTM, and LSTM-CNN

Zainab Alshingiti [1], Rabeah Alaqel [1], Jalal Al-Muhtadi [1,2], Qazi Emad Ul Haq [3,*], Kashif Saleem [2] and Muhammad Hamza Faheem [3]

[1] Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11653, Saudi Arabia
[2] Center of Excellence in Information Assurance (CoEIA), King Saud University, Riyadh 11653, Saudi Arabia
[3] Center of Excellence in Cybercrimes and Digital Forensics (CoECDF), Naif Arab University for Security Sciences (NAUSS), Riyadh 11452, Saudi Arabia
* Correspondence: qabdulrab@nauss.edu.sa

Abstract: In terms of the Internet and communication, security is the fundamental challenging aspect. There are numerous ways to harm the security of internet users; the most common is phishing, which is a type of attack that aims to steal or misuse a user's personal information, including account information, identity, passwords, and credit card details. Phishers gather information about the users through mimicking original websites that are indistinguishable to the eye. Sensitive information about the users may be accessed and they might be subject to financial harm or identity theft. Therefore, there is a strong need to develop a system that efficiently detects phishing websites. Three distinct deep learning-based techniques are proposed in this paper to identify phishing websites, including long short-term memory (LSTM) and convolutional neural network (CNN) for comparison, and lastly an LSTM–CNN-based approach. Experimental findings demonstrate the accuracy of the suggested techniques, i.e., 99.2%, 97.6%, and 96.8% for CNN, LSTM–CNN, and LSTM, respectively. The proposed phishing detection method demonstrated by the CNN-based system is superior.

Keywords: phishing detection; website URL; deep learning; convolutional neural network (CNN); LSTM; cyber-attack detection

Citation: Alshingiti, Z.; Alaqel, R.; Al-Muhtadi, J.; Haq, Q.E.U.; Saleem, K.; Faheem, M.H. A Deep Learning-Based Phishing Detection System Using CNN, LSTM, and LSTM-CNN. *Electronics* **2023**, *12*, 232. https://doi.org/10.3390/electronics12010232

Academic Editors: Enzo Pasquale Scilingo and Dah-Jye Lee

Received: 13 October 2022
Revised: 25 November 2022
Accepted: 29 November 2022
Published: 3 January 2023

1. Introduction

Life has become faster and more accessible because of the evolution of communication technologies and digitalization, especially during the lockdown due to the COVID-19 pandemic, when all transactions and life needs needed to procured online, i.e., shopping and transactions, as compared to doing so physically. To fulfil daily needs on online systems, you can simply open your smart device, and search for the website as you want, such as a pharmacy, shopping store, learning platform, or bookstore. On the other hand, the growth of E-services expands attackers' opportunities to gain or misuse users' information such as their names, phone numbers, identification, and credit card information. As a result, users face a variety of online threats and cyber-attacks every day. Phishing has different types, it could be via electronic mail (E-mail), SMS (Short Message Service), or URL (Uniform Resource Locator), to name a few. Phishing can compromise all types of data sources including personal information and online accounts, and gain access and modification to connected systems [1].

In some cases, hackers stop phishing when they steal enough information for financial gain while other hackers seek to earn more information by logging into specific companies to make more malicious attacks against their employees. Consequently, hackers use different and new techniques to fool users such as sending URLs that look like a website for banking or shopping; at the time when the user opens the URL and conducts transactions,

the hacker is capable of stealing a lot of important data such as account details, credit card information, users' personal information, passwords, and identity [2].

URL phishing is a cyber-attack that uses URLs and e-mails as a technique to trick users into believing that the URL or e-mail is a trustworthy mechanism in electronic communication, such as a note from their company or a request from their bank, for instance, to download the attachment or to click a link. At that moment, attackers are able to access the user's data. Furthermore, phishing websites or e-mails are designed to mimic the look of a real company webpage/email [3].

The rapid evolution of intelligent techniques such as machine learning (ML) and deep learning (DL), which fall under artificial intelligence (AI), are effective in providing security for the operations of computing and cybersecurity management. The variety of AI characteristics, from detecting and extrapolating patterns, to providing security to adapt to a new environment make it a pivotal part of technological systems such as computer vision and cybersecurity [4].

To perform feature extraction and selection in classic machine learning techniques, human expertise is needed. Feature selection and classification tasks are separated. In order to optimize the models' performance, deep learning fills that gap using a single phase for detection and classification. Due to automatic learning and feature extraction, deep learning models minimize the need for manual feature engineering and reliance on third-party services, unlike machine learning [5,6]. Moreover, high performance and end-to-end problem-solving are the major advantages of deep learning over traditional machine learning techniques, especially in cases of large datasets such as speech recognition, image classification, and detection of phishing [7–10]. Bagui et al. [11] conducted a comparison of ML and DL models in different studies and the authors concluded that DL models performed better for detecting phishing websites than ML models in terms of accuracy.

Selecting the best method for a given application is not simple. The accuracy and efficiency of the model would eventually suffer if the wrong algorithm or method were used [12], especially given how frequently phishers alter their attack strategies to take advantage of weak points in systems and users' ignorance. Numerous anti-phishing technologies have been developed as a result to identify phishing risks early and shield users from such attacks. Security methods based on deep learning mechanisms are being employed more frequently across a variety of industries to combat emerging phishing assaults [13,14].

Deep learning applications are used in different industries such as autonomous driving, facial recognition, and medical devices, to name a few. Deep learning trains machines to mimic human brains through learning by example. Furthermore, through the process of "deep learning", a computer model can directly learn how to execute classification tasks from large datasets that include text, sound, and images. Deep learning models can attain better results; sometimes the results even exceed human performance. For training deep learning models, a large amount of labeled data is required, substantial computing power, and neural network architectures that contain numerous layers [15,16].

The robustness of deep learning algorithms has encouraged researchers to propose many methods for dealing with phishing websites by extracting features for classifying URLs. Numerous methods that assist in detecting phishing attacks have been applied by using different, new, and known features such as URL length, frequency of keywords, lexical features, and by incorporating new features.

LSTM (long short-term memory) is a form of recurrent neural network (RNN) that gains superior results when dealing with time-series data, removing vanishing gradients and long-term dependencies. The architecture of LSTM is made up of a cell and three gates (input, output, and forget) [17,18] as shown in Figure 1.

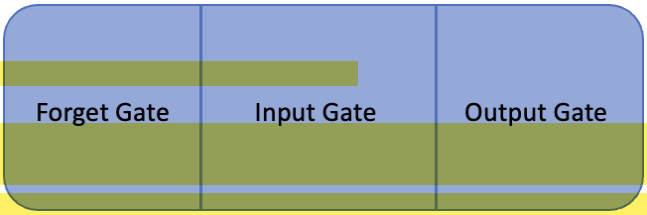

Figure 1. LSTM basic architecture.

A convolutional neural network is a kind of neural network that requires large, labeled data for training. CNNs play a significant role in many problems such as image classification, object recognition, phishing detection, and diagnosis of medical diseases. Input, convolution, pooling, and fully connected layers are the main layers needed to construct a CNN as shown in Figure 2. Accelerating the learning process has led CNN to accomplish great and high results for many problems [17].

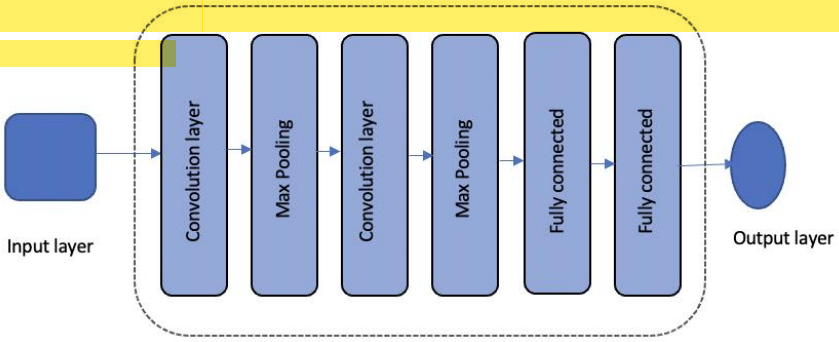

Figure 2. CNN basic architecture.

LSTM–CNN architecture involves both CNN and LSTM methods as shown in Figure 3 in order to make use of the benefits of both methods and accomplish excellent performance. Since CNN and LSTM show high performance in overcoming classification, detection, and recognition tasks [17], to using these three methods for the phishing detection task is promising.

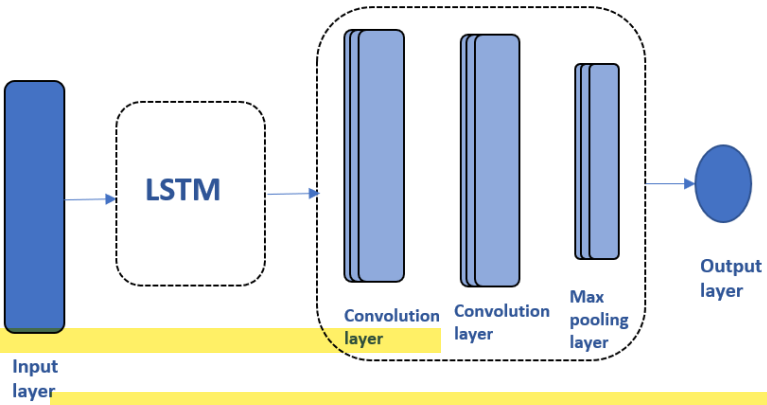

Figure 3. The architecture of LSTM–CNN.

As a result, we were motivated to find a solution to phishing websites effectively using deep learning. This paper used an empirical method to investigate the performance of the three techniques LSTM, CNN, and LSTM–CNN in order to produce great results in phishing detection. The goal of this paper was to classify whether the URL was phished or legitimate by using LSTM, CNN, and LSTM–CNN.

To determine if the URLs are phished or legitimate, we suggest a phishing detection system based on deep learning techniques. The suggested approach is useful for deep

learning-based detection and classification systems in the fields of information security and cybersecurity. In order to classify phishing URLs and stop financial losses and cybercrimes, our work offers a great contribution to the efficacy of using LSTM, CNN, and LSTM–CNN.

The following points state the contribution of the proposed work:

- An examination of the methods currently used to identify phishing websites.
- Analysis and use of three state-of-the-art deep learning methods, LSTM, CNN, and LSTM–CNN, to predict phishing URLs.
- Presentation of an efficient deep learning architecture based on CNN due to its capacity to identify patterns, extract features, and automatic and accurate classification of URLs.
- Comparison and evaluation of suggested LSTM, CNN, and LSTM–CNN models.
- Consideration of a dataset with 30 features after a feature selection process.
- Highlighting several restrictions based on the conclusions of earlier investigations and suggestion of potential fixes for these issues.

The remainder of this paper is structured as follows: A literature review is presented in Section 2. Section 3 discusses our proposed solution along with its methodology. Section 4 contains experimental results and a discussion. Section 5 is a comparison of existing works. Section 6 focuses on the conclusion and future work.

2. Literature Review

The phishing website problem is complex and is a challenge in itself, because no definitive solution exists to put an end to all the threats effectively. To identify phishing websites, deep learning-based phishing website detection solutions have arisen. Moreover, deep learning has become more promising in cyber security. In this section, several previous works that use deep learning approaches for phishing website detection are shown in Table 1.

2.1. Long Short-Term Memory (LSTM)

Yang et al. [19] presented a new method that uses the LSTM and recurrent neural network (RNN) algorithms for detecting phishing attacks that adopts the LSTM deep learning method and optimizes the training of the model with the combined characteristics of RNN. The main advantages of using LSTM are its ability to incorporate large volumes of data and capacity to automatically learn complex features. This solves a complex problem for other machine learning methods. The datasets used were from yahoo and PhishTank. This work showed an accuracy of 99.1%.

2.2. Convolutional Neural Network (CNN)

A model based on deep learning proposed in [20] utilized a character-level CNN to detect phishing URLs. The study implemented a system of phishing detection by using CNN at a character level to learn the URL's sequential information, then max-pooling was applied to determine important features, which were then fed to fully connected layers for classification. To train the network, the stochastic gradient descent algorithm (SGD) was used. The results show that the suggested model attained an accuracy of 95.02% on the given dataset. Furthermore, the model's accuracy on benchmark datasets was 98.58%, 95.46%, and 95.22%, which performed better than the current phishing URL models compared to the various machine and deep learning algorithms.

Shweta et al. [21] presented a phishing detection system using deep learning techniques to prevent phishing attacks. The dataset contained 37,175 phishing URLs and 36,400 legitimate ones. The study was conducted by applying CNN. The advantage of this system is that no feature engineering is required since the CNN extracts features from the URLs automatically through its hidden layers. The framework consists of the input text being passed through the embedding layer, and a matrix being created and passed to CNN. The accuracy of the proposed system achieved was 98.00%.

Table 1. Comparative analysis of literature review.

Ref.	Brief Problem Statement	Methodology Based on DL	Feature Extraction Method	Classification Method	Dataset	Training Instances	Testing Instances	Performance Measures
[19]	A new detection system for phishing websites using LSTM and RNN.	LSTM Keras and RNN	-	LSTM with sigmoid	Yahoo Directory and PhishTank	70%	30%	Accuracy
[20]	Detect phishing without requiring essential manual feature engineering or prior knowledge about phishing.	A fast deep learning-based solution model	The features extracted from the URL do not need manually designed hand-crafted features; it is independent of network access	Naïve Bayes, logistic regression, random forest, XGBoost, and deep neural networks	PhishTank, Common Crawl, and Alexa	Randomly split		Accuracy, recall, F1-Score, precision, area under the curve (AUC), training, and test time
[21]	Design phishing detection system.	CNN	CNN	CNN	High-risk URLs	Phishing and legitimate URLs		F-1 score precision, recall, and accuracy
[22]	In response to the new threat called two-dimensional code phishing attacks.	Improved faster R-CNN	The heuristic-based approach	SVM	FlickrLogos-32	10 training images	30 verification images, and 30 test images	Precision, recall, and F1-measure
[23]	Proposed techniques to fight phishing attacks.	Deep learning-based data-driven end-to-end automatic phishing webpage classification	Learns context features from HTML documents without requiring extensive manual feature engineering	CNNs	HTML documents using a web crawler	23,000 legitimate URLs and 2300 phishing URLs		The precision, true positive rate, F-1 score metrics, AUC, and the receiver operating characteristic (ROC)
[24]	A phishing detection system using CNN with n-gram features that are extracted from URLs.	Deep learning n-gram method, CNN model	N-gram	Deep learning-based	High-risk URLs	85%	15%	Accuracy and run-time efficiency

Table 1. *Cont.*

Ref.	Brief Problem Statement	Methodology Based on DL	Feature Extraction Method	Classification Method	Dataset	Training Instances	Testing Instances	Performance Measures
[25]	A novel deep learning architecture, Texception, that predicts phishing attack.	CNN, URLNet model, and LR		Binary Cross entropy loss function along with SGD optimize	Microsoft SmartScreen service, Microsoft's anonymized browsing telemetry data	The first two weeks		Accuracy
[26]	A deep learning-based method for very accurate phishing site identification.	CNN	CNN	CNN	UCI dataset	90%	10%	Accuracy
[27]	Analyze the performance of various deep learning algorithms in detecting phishing activities.	DNN, CNN, LSTM, and GRU	Traditional neural network	CNN	UCI dataset	80%	20%	Confusion matrix, ACC, PR, RC, F1, other metrics, including FPR, FNR, and AUC.
[28]	A deep learning-based phishing detection solution that leverages URL and website content.	CNN and LSTM algorithm	Feature extractor algorithm	IPDS (CNN + LSTM) knowledge model	PhishTank and Common Crawl	70%	30%	Classifier prediction performance, training time, and accuracy
[29]	Present PhishTrim, a quick and simple deep representation learning-based phishing URL detection technique.	Skip-gram pre-training model, Bi-LSTM, and CNN	CNN	Multiple convolution structures	PhishTrim			Accuracy
[30]	An anti-phishing system to protect users against phishing.	LSTM, CNN, and RNN		LSTM cross-entropy loss function	PhishTank, VirusTotal and using Yandex search API.	80%	20%	Accuracy and precision

A relative detection method was suggested in [22], which allowed for the identification of a two-dimensional code phishing attempt. Information was gathered from the FlickrLogos-32 dataset, a publicly accessible logo dataset with 32 unique logo brands. The study was conducted by enhancing the traditional approach, which is an improved feature pyramid network (FPN) combined with a faster R-CNN logo identification technique. The three logo processes were the main processes of the system, which are extraction, recognition, and identification. Extracting logo images from two-dimensional code is known as logo extraction. Based on the retrieved logos, the identification and recognition of the logos were performed using faster R-CNN. The final step in the identification process involves assessing the logo's consistency between the actually identified object and its described identity. In comparison to other logo recognition methods and phishing detection methods, the findings demonstrated the method's effectiveness in logo recognition, which may be used for two-dimensional code phishing assault detection.

HTMLPhish is a deep learning-based platform that relies on data-driven end-to-end automatic phishing web page classification, as proposed by Chidimma et al. [23]. The dataset includes more than 50,000 HTML documents and a full dataset of HTML contents was presented in a real-world distribution. The data were acquired from HTML documents using a web crawler. HTMLPhish employed CNNs to learn the semantic dependencies in the textual contents of HTML documents in order to learn the relevant feature representations. Additionally, they used convolutions on a combination of the character and word embedding matrix to ensure that new words were effectively incorporated into the test HTML documents. Without taking into account intensive manual feature engineering, this technique could analyze context features from HTML pages. The results showed that HTMLPhish obtained over 93% accuracy, which indicates good result.

Due to internet users' exposure to cyber threats and security flaws, artificial intelligence-based algorithms through machine learning and deep learning techniques were developed [24]. The authors aimed to construct a system that detects phishing to overcome cyberattacks using a CNN with n-gram features. The system extracts these features from URLs, determining which n-gram feature extraction technique is more effective and which parameter works best. The best results are achieved with single characters. Using 70 characters in model training gives 34 s for training one epoch and 0.008 s for URL classification. With the high-risk URL dataset, reaching an accuracy of around 88.90% is excellent.

Texception is a new deep learning architecture [25] that predicts whether the input URL is a phishing link or not. Texception is different from classical approaches since it uses two levels of information from the URL, which are character-level and word-level, depending less on manually crafted features. Texception grows wider or deeper through different parallel convolutional layers. For new URLs using the Microsoft SmartScreen service dataset, Texception generalizes better. The results of production data showed that Texception achieved magnificent performance. The true positive rate increased by 126.7% with a (0.01%) false-positive rate.

The improvement of cyber defense and effective phishing detection is required to cope with the increased exposure to various cyberattacks owing to the faster growth of phishing websites. Yerima et al. [26] used a 1D CNN-based model that utilizes CNN for its capability in differentiating sites of legitimate or phishing. According to the authors, the model evaluated a website dataset including 4898 and 6157 phishing and legitimate websites, respectively. The model is used to detect unseen phishing websites. Furthermore, the model gained 98.2% and 0.976 as a phishing detection rate and F1-score, respectively.

2.3. Integration of LSTM and CNN

Quang et al. [27] concentrated on analyzing the performance of different deep learning algorithms in detecting phishing websites to aid organizations in choosing and adopting suitable solutions based on their technological needs. The data contains 11,055 phishing and benign URLs. They utilized various deep learning algorithms, which comprised DNN, CNN, gated recurrent unit (GRU), and LSTM. In order to find the optimal parameter to

achieve good accuracy, the model was tested on different architectures for each of the deep learning algorithms. The results demonstrated that a deep learning algorithm gains the best measure of overall performance metrics.

Image classification and natural language can both benefit from deep learning approaches. Adebowale et al. [28] proposed an intelligent phishing detection system (IPDS) to explore the potential of distinguishing phishing URLs from unique legitimate URLs. IPDS builds a hybrid classification model using LSTM and CNN. Around one million legitimate and phishing URLs were used on the dataset collected from PhishTank and Common Crawl. To build the IPDS, the LSTM and CNN classifier used over 10,000 images and one million URLs for training. The sensitivity of IPDS was determined by several factors such as split issues, number of misclassifications, and the type of feature. IPDS achieved 93.28% as the accuracy of classification.

The detection rules of many phishing detection techniques are difficult to update in response to changes in attack trends and computationally expensive. PhishTrim was proposed by Zhang et al. [29], which is a lightweight phishing URL detection method based on deep representation learning. The skip-gram pretraining model was used to obtain the URLs' initial embedding representation. Furthermore, to extract context dependency and learn the deep representation of URLs, Bi-LSTM was used. the local n-gram features were extracted via CNN, and the PhishTrim dataset was used.

As a result of the increase in electronic shopping (e-shopping) and electronic banking (e-banking), hackers can steal users' personal information and critical details through different ways by passing themselves off as trusted websites. To protect users from such cases, Yazhmozhi et al. [30] proposed an anti-phishing system based on LSTM and CNN. The dataset comprised nearly 200,000 URLs taken from PhishTank, VirusTotal, and by using Yandex search API. The proposed system performs well, with 97% precision and 96% accuracy. The model can be used in web browsers since it is deployed with a simple UI.

After a comprehensive literature review, phishing detection research is a challenging task, since phishers are rapidly developing efficient ways to bypass the current detectors. Research on phishing detection approaches can be categorized depending on their input such as URL, email, visual screenshot, logos, and HTML content. In terms of URL as input, most of the studies have proven that URL features such as URL length, characters, frequency of keywords, and frequency of auspicious symbols signify well on the datasets collected from VirusTotal, PhishTank, OpenPish, and other open phishing platforms. The results of these studies showed accuracy reaching 90% and more using deep learning-based methodologies, mainly DNN, CNN, and LSTM. On the other hand, some studies use small datasets, which affect the accuracy of the proposed systems. Furthermore, some studies used the same deep learning method for feature extraction and classification obtaining different accuracies; in addition, the training time was long. Hence, there is a need for a system that can help detect phishing URLs efficiently and effectively. Deep learning has attracted increased interest recently due to its performance and ability to learn the features instantaneously without any manual feature engineering. Under those promises, we used deep learning to detect phishing URLs using LSTM, CNN, and LSTM–CNN to show their performances in detecting phishing URLs. To the best of our knowledge, no previous work uses the three DL methods and compares their results. The dataset used in this work contains 20,000 URLs including 9800 phishing ones [31]. The primary difference of our approach with regard to the previously cited deep learning-based ones is that we extracted the most discriminative features for the dataset and proposed the use of a light-weight CNN-based model for the accurate detection of phishing websites, which turned out to be conducive to the improvement of phishing detection performance.

3. Methodology

Detecting phishing URLs is an important aspect of cybersecurity. Commonly, many phishing URLs appear as legitimate URLs to the users because of the complex formulation of URLs by attackers. As a result, attackers can gain access to the personal information

of users, which can be misused. This paper proposed a phishing detection system for detecting phishing URLs. In order to detect phishing URLs and show the robustness of the system, the system was implemented by using two different techniques. The following sections describe the methodology used, dataset preparation, deep learning approaches, and the model's training and testing detail.

3.1. Proposed System

In this section, the important details of the models' configuration are discussed. The framework of the model incorporates of four stages as shown in Figure 4. The first stage concerns the features of the URLs, which are obtained from the dataset [31]; the second stage involves pre-processing, in which we detected null values and scaling values of feature selection, which contributes most to the target variable by using SelectKBest; the third stage is the training of three different models, namely LSTM, CNN, and LSTM–CNN by building a deep learning approach. Finally, as the evaluation of the approach using a number of indicators to measure how the model performs in detecting phishing websites, the fourth stage is the classification of the webpage URLs as legitimate or phishing.

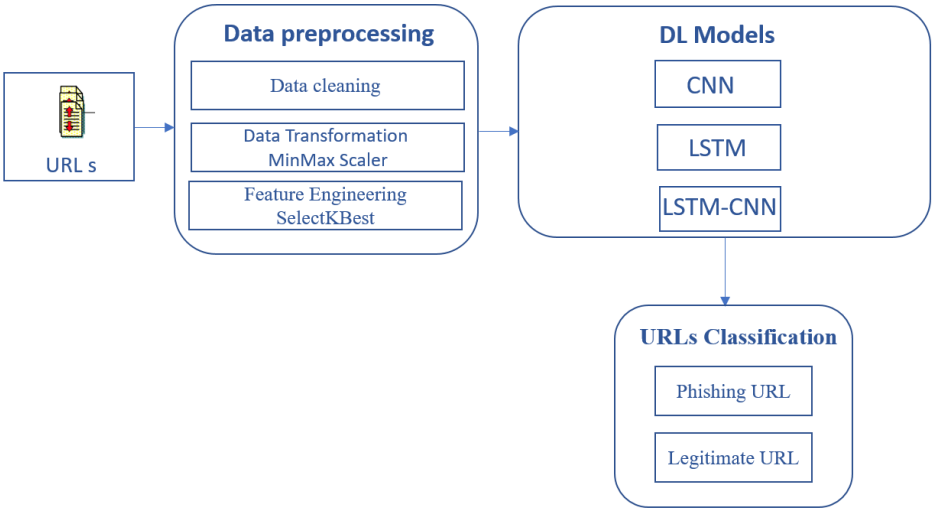

Figure 4. Framework of the proposed system.

3.2. Dataset Preparation and Preprocessing

Data collection plays an essential role in terms of research validity and reliability. In our approach, we made use of appropriate and consistent data, so the system's training is robust. After prepossessing the dataset containing the URL features, with 20,000 records of 80 features, there were a lot of features in the dataset; therefore, the SelectKBest method was used with the value of the 30 best features. The dataset under consideration was processed in the data preprocessing stage, which included detecting null values in addition to scaling each feature to a given range using the MinMaxScaler method. The obtained dataset after preprocessing was individually taken into account during various experiments over the LSTM, CNN, and LSTM–CNN.

3.3. Training and Testing

The dataset was divided into 20% as testing and 80% as training. The distribution of training and testing sets is shown in Table 2. One of the aspects affecting the effectiveness of deep learning algorithms is the selection of hyperparameters during training. Hyperparameter values can be optimized to improve the accuracy of phishing website detection models. These parameters comprise the number of layers, the number of neurons in each layer, the batch size, the learning rate, the dropout rate, the number of epochs, the type of activation function, the type of optimizer, the learning rate, and the dropout rate [32]. Choosing an appropriate number of parameters enhances the LSTM, CNN, and

LSTM–CNN models' performance, so each parameter was selected based on the value that enhanced performance. One of the main parameters of the system is the age, which is considered as the number of iterations of training after the deep learning model is built and compiled, its value set to 50 epochs. The parameters are stated in Table 3.

Table 2. Training and testing dataset distribution.

Dataset Distribution	Phishing	Benign
URLs for training	7840	8160
URLs for testing	1960	2040

Table 3. Parameters.

Parameters	Values
Activation Function	Relu
Epochs	50
Batch size	1200
Optimizer	Adam
Dropout	0.2

3.4. Deep Learning Approaches

Deep learning, a subfield of machine learning, has gained great attention over the previous decade. Recent advances in processing power and increased data storage capacities have greatly aided the ability to apply deep learning approaches. Deep learning models have produced excellent results using large datasets for a variety of challenges, including image processing, natural language processing and machine translation. Moreover, the challenge of phishing URL classification has also been undertaken using deep learning systems, with encouraging results [33]. Different classification methods are applied to detect phishing websites and then evaluated by different performance metrics. The models examined in this study are LSTM, CNN, and LSTM–CNN. Convolutional layers are defined by their ability to learn internal representations and retrieve meaningful knowledge of data; LSTM networks, on the other hand, are efficient at detecting both short- and long-term dependencies. Based on the experimental results, the CNN model shows great results in terms of performance. Furthermore, we explain each of the three models below.

- Long short-term memory (LSTM): Long short-term memory is an adaptive recurrent neural network (RNN), which is a type of recurrent neural network in which a memory cell, in addition to the conservative neuron, switches each neuron on account of an internal state. The layers of LSTM comprise memory blocks, which repeatedly link blocks; one or more memory cells with recurrent connections can be found in each block.
  As a result, a typical LSTM cell has an input gate that controls data input from outside the cell and determines whether the data in the internal state is kept or overlooked, as well as an output gate that prohibits or enables the inner state's ability to be viewed from the outside [34]. LSTM has been shown to be an effective strategy for detecting phishing URLs [35,36]. The workflow of LSTM for classifying a URL starts after loading, preprocessing, and splitting the dataset. The LSTM model starts with the first layer, which is the input layer that uses a 79-length vector, and then the LSTM layer, which includes 128 neurons and acts as the model's memory subset. Following LSTM, the dense layer—an output layer with a sigmoid function—assists in providing the labels.
- Convolutional neural network (CNN): CNN is a discriminative architecture that works effectively at processing grid-based two-dimensional data, including images and videos. In terms of time delay, the CNN outperforms the neural network (NN). The weights are shared in a temporal dimension in the CNN, which reduces calculation time. The

standard NN's generic matrix multiplication is thus replaced in the CNN. As a result, the CNN technique minimizes the weights, lowering the network's complexity [34]. The workflow of the CNN for classifying a URL starts with the first step by fetching the labeled training data of the URLs, then divides into train and test sets at random. After we prepared the training and test data, the data was finally trained by creating the architecture of the CNN including the input, output, and layers. After each convolution, we incorporated a max-pool layer to capture the essential elements from each convolution and convert them into a feature vector. Next, we added dropout regularization to ensure that that model did not overfit. The model classifies the output produced by this layer when a sigmoid function is used.

- LSTM—CNN: The model consists of CNN layers that extract features from input data and LSTM layers that predict sequences [37]. Furthermore, a study [38] found that combining a 1D convolution layer and an LSTM layer improves the accuracy of malicious URL identification when compared to models that exclusively use LSTM layers. As a result, when constructing the system, we chose 1D CNN and LSTM architecture to train the URL features.

The workflow of CNN–LSTM as shown in Figure 3; after preprocessing the dataset, it splits into train and test sets, followed by data normalization before feeding into the model; lastly, the model is passed to the CNN and LSTM layers, in addition to the dense layer to avoid overfitting of the dataset, and finally, the model classifies the results of the output produced by this layer when a sigmoid function is used.

4. Evaluation and Results

This section evaluates the proposed system and presents the results.

4.1. Evaluation Metrics

This section summarizes the metrics used to measure the results of the deep learning approaches. Generally, using results of the classification algorithm, the performance of machine learning prediction algorithms are evaluated. In this study, the prediction outcomes were examined using metrics including precision, recall confusion matrix, and accuracy of the system to estimate the system [39].

Precision: The precision of the prediction algorithm is the number of phishing webpages correctly classified as actual phishing webpages.

$$Precision = \frac{TP}{TP + FP} = \frac{True\ positive}{Total\ predicted\ positive} \quad (1)$$

Recall: Recall of the prediction algorithm is the number of correct phishing URL predictions made over all URLs in the dataset.

$$Recall = \frac{TP}{TP + FN} = \frac{True\ positive}{Total\ predicted\ positive} \quad (2)$$

Accuracy: The accuracy of the prediction algorithm is the ratio of the total number of correct predictions of class to the actual class of the dataset. Equation (3) calculates the accuracy of the model. Typically, any prediction model produces four different results, namely true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (3)$$

F1-Score: The process of taking the harmonic mean of a classifier's precision and recall. It can be combined into a single metric.

$$F1 - score = \frac{2 \times (Precision\ \times Recall)}{(Precision + Recall)} \quad (4)$$

4.2. Results

For the experimental results, we calculated the accuracy, precision, recall and F1 score of the prediction algorithms. In the majority of prediction models the proposed system was evaluated based on the accuracy of the prediction model, which has been identified as one of the common performance measures. The prediction accuracy of the approaches presented in this paper can be found in Section 3. We used a dataset that consists of 20,000 records of URLs consisting of 80 features. In the preprocessing stage we detected null values and scaled features, and then selecting 30 features using SelectKBest, we trained the LSTM, CNN, and LSTM–CNN classifiers based on these features.

The three proposed methods showed good results, which are shown in Table 4, also reflecting the optimal choice of parameters. After implementing, training, and testing the LSTM, CNN, and LSTM–CNN techniques, the results showed some level of improvement in phishing detection through the CNN algorithm, since it had the highest accuracy at 99.2%, followed by the LSTM–CNN algorithm, which achieved 97.6%, while LSTM achieved 96.8% prediction accuracy as illustrated in Figure 5. Because CNN outperforms the other two models in terms of accuracy and other performance metrics, it is superior to them due to different reasons: First, CNN can perform well on text classification problems while LSTM performs for sequential data, since LSTM can learn the texts and the relation between the tokens very well. Moreover, CNN takes less time and is more effective than the LSTM-based approach. In addition, it requires fewer parameters for training compared to LSTM, which reduces the complexity of the model. Additionally, CNN runs one order of magnitude faster than both LSTM and LSTM–CNN. Finally, the computations in CNNs can occur in parallel, in contrast to LSTM, which captures the dependency across time sequences in the input vector.

Table 4. The performance results.

Evaluation Metric	LSTM (%)	CNN (%)	LSTM-CNN (%)
Accuracy	96.8	**99.2**	97.6
Precision	95.9	99	96.9
Recall	97.5	99.2	98.2
F1-score	96.8	99.2	97.6

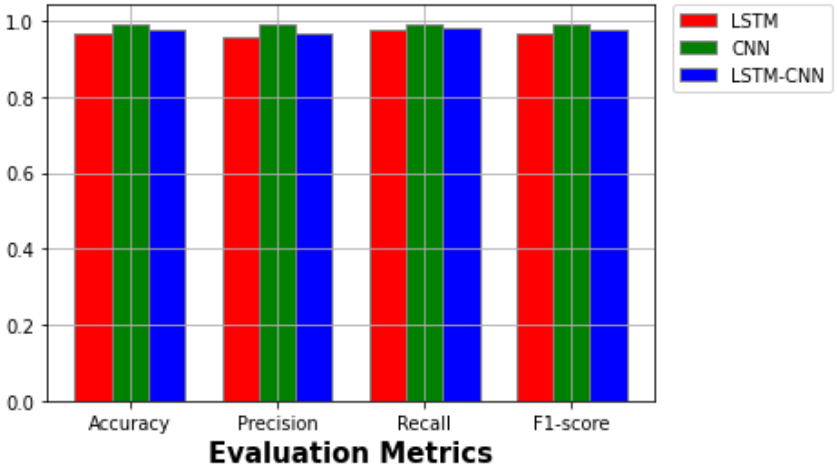

Figure 5. Evaluation metrics.

For the LSTM, in Figure 6, the confusion matrix of the LSTM model is shown. The percentage of predicted values is shown on the x-axis, and the percentage of true values is shown on the y-axis. It is obvious that the LSTM algorithm predicted 1912 (true positive) samples correctly, with 80 (false positive) misclassifications.

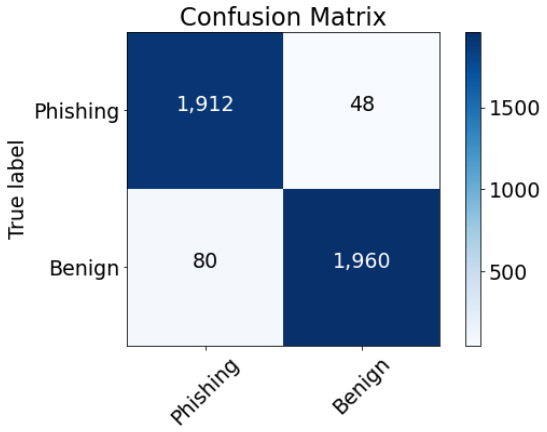

Figure 6. Confusion matrix of LSTM.

The confusion matrix of the CNN algorithm is illustrated in Figure 7. The percentage of predicted values is shown on the x-axis, and the percentage of true values is shown on the y-axis. It is obvious that the LSTM algorithm predicted 1946 (true positive) samples correctly, with 18 (false positive) misclassifications. Figure 8 illustrates the confusion matrix of the LSTM–CNN algorithm. The percentage of predicted values is shown on the x-axis, and the percentage of true values is shown on the y-axis. It can be seen that the LSTM–CNN algorithm predicted 1925 (true positive) samples correctly with 60 (false positive) misclassifications.

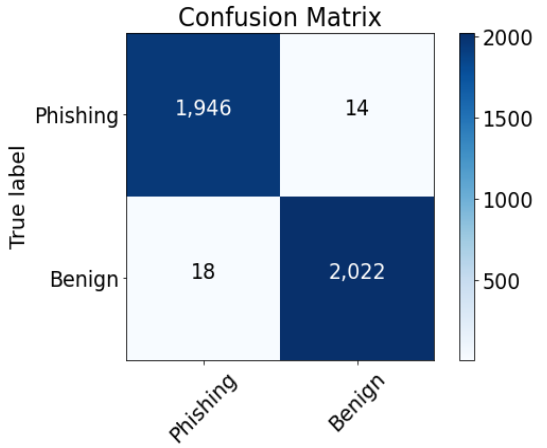

Figure 7. Confusion matrix of CNN.

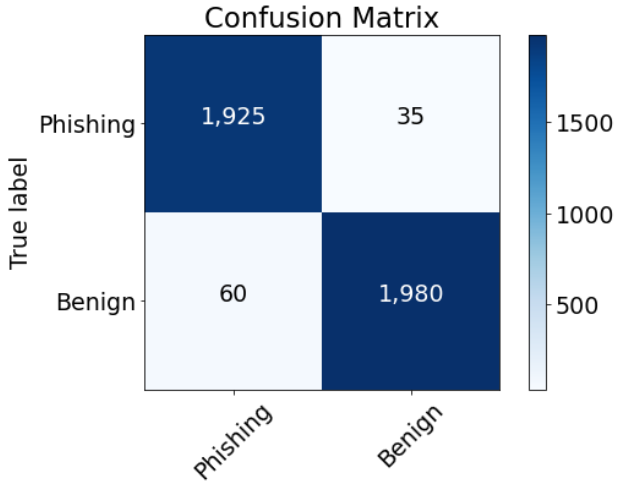

Figure 8. Confusion matrix of LSTM–CNN.

After analyzing by considering the outcome, we could say that the CNN algorithm outperforms the LSTM–CNN and CNN algorithms in the detection of phishing.

5. Comparison with Existing Approaches

It is important to shed light on previous works that have used similar approaches and methodology to our work. The proposed CNN architecture provides excellent results compared to LSTM–CNN and LSTM. Furthermore, we also compare our proposed model with already existing techniques that have used CNN and LSTM in Table 5. The comparison is based on the proposed methodology, the data set used, advantages and disadvantages, and the system accuracy of the existing works.

Table 5. Comparison of existing approaches.

Ref.	Proposed Methodology Based on DL	Dataset	Advantages/Disadvantages	Accuracy
[27]	DNN, CNN, LSTM, and GRU	(UCI)	**Advantages** • Secure connection between a mail user agent and a mail transfer agent. • Fast classification process. **Disadvantages** • It takes more time for parameter selection and network learning.	96.70%
[28]	LSTM and CNN hybrid model	PhishTank and Common Crawl datasets	**Disadvantages** • The models need more computing power.	93.28%
[30]	LSTM and CNN	PhishTank, VirusTotal,	**Advantages** • High precision and less computationally expensive. **Disadvantages** • Insufficient parameter selection techniques. • Parameter tuning is performed manually.	96%
[39]	CNN–LSTM	PhishTank and OpenPhish	**Advantages** • The method offers a diverse combination of CNN and RNN. **Disadvantages** • Increases the amount of time needed to train the model.	98%
[40]	CNN–LSTM	PhishTank	**Advantages** • The experimental results showed that this algorithm is more accurate than traditional algorithms. **Disadvantages** • Improvement needed, such as performing multiple classifications.	98.18%
[41]	CNN is used to detect phishing URL features based on CNN and Bi-LSTM.	PhishTank, MalwarePatrol, DMOZ and Alexa	**Disadvantages** • There is no standard and holistic guideline for selecting these hyperparameters to achieve the highest performance.	96%

Table 5. *Cont.*

Ref.	Proposed Methodology Based on DL	Dataset	Advantages/Disadvantages	Accuracy
[42]	CNN	ISCX-URL2016	**Disadvantages** • The proposed approach is relevant only to URL characteristics of the same dataset.	99%
[42]	LR, SVM, RF, RNN, RNN-GRU, and RNN–LSTM.	ISCX-URL2016	**Advantages** • The proposed solution is used in live web browsing sessions in a real-time environment. **Disadvantages** • Part of the information is lost for long URLs with more than 200 characters. • The system requires more computational power and is time intensive.	99%
[43]	CNN, CAE	ISCX-URL2016	**Advantages** • Integrates a convolutional autoencoder to rebuild a URL and calculate the abnormal score for a phishing attack. **Disadvantages** • Optimized to character-level features among the numerous features that affect URLs.	88%
Proposed	**CNN**	**ISCX-URL2016**		**99.2%**

Limitations

After testing and evaluating our proposed system, we can see that the system outperforms existing methodologies and showed excellent results. However, the proposed system has some shortcomings. The model does not check the status of the URL of the website, i.e., whether the website is active or not, which impacts the results. To overcome this limitation, it might be necessary to speed up the training process and improve feature engineering, which would then allow us to verify the website's state and improve training process accuracy.

6. Conclusions and Future Work

The improvement of technologies has had a significant impact on increasing online purchases and transactions, which make our day-to-day tasks easier. On the other hand, online transactions lead to unauthorized access to the sensitive information of users, individuals, or enterprises. Security is the most important aspect of protecting users from phishers who steal information while they are communicating through internet applications. Phishing is one of the known attacks that gain users' information through a URL that looks identical to the actual webpage. Detecting phishing attacks plays a significant role in preventing attackers from gaining access to users' information. As there is a growth in the number of victims owing mainly to inefficient security technology adoption, an intelligent technique is needed to protect users from cyber-attacks. With the rapid development of deep learning techniques, deep learning has proven a valuable development compared to traditional signature-based and classic machine learning-based solutions due to its high performance and end-to-end problem-solving. In this work, the LSTM, CNN, and LSTM–CNN algorithms were proposed to detect and classify the URLs of the websites as either

phishing or legitimate. Based on the evaluation of the proposed system, the detection of phishing websites accomplished excellent results. The proposed deep learning algorithms applied to the same dataset varied in their performance. The CNN algorithm outperformed LSTM–CNN and LSTM in terms of accuracy, which reached 99.2%, while LSTM–CNN and LSTM achieved accuracies of 97.6%, and 96.8%, respectively. In the future, we aim to enhance the training process by reducing training time and improving feature engineering in order to verify websites' states and improve the training processes' overall accuracy. Furthermore, we also intend to present an approach that considers the webpage context as well as the URL in order to detect phishing websites.

Author Contributions: Conceptualization, Q.E.U.H. and J.A.-M.; data curation, Z.A., R.A. and Q.E.U.H.; formal analysis, Z.A., R.A., J.A.-M., Q.E.U.H. and K.S.; funding acquisition, J.A.-M.; methodology, Z.A., R.A., J.A.-M., Q.E.U.H. and K.S.; project administration, J.A.-M. and K.S.; resources, J.A.-M.; software, Z.A., R.A. and Q.E.U.H.; supervision, J.A.-M., Q.E.U.H. and K.S.; validation, Z.A., R.A., J.A.-M., Q.E.U.H., K.S. and M.H.F.; visualization, Z.A., R.A., J.A.-M., Q.E.U.H., K.S. and M.H.F.; writing—original draft, Z.A. and R.A.; writing—review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

Funding: The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through project number IFKSURG-2-110.

Acknowledgments: The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through project number IFKSURG-2-110.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. What Is Phishing? Phishing. Available online: https://www.phishing.org/what-is-phishing (accessed on 28 October 2022).
2. What Is Phishing: Attack Techniques & Scam Examples: Imperva (2020) Learning Center. Available online: https://www.imperva.com/learn/application-security/phishing-attack-scam/ (accessed on 28 October 2022).
3. Phishing | Phishing Techniques. Phishing.org. 2022. Available online: https://www.phishing.org/phishing-techniques (accessed on 21 April 2022).
4. Basit, A.; Zafar, M.; Liu, X.; Javed, A.R.; Jalil, Z.; Kifayat, K. A comprehensive survey of AI-enabled phishing attacks detection techniques. *Telecommun. Syst.* **2021**, *76*, 139–154. [CrossRef] [PubMed]
5. Alsariera, Y.A.; Elijah, A.V.; Balogun, A.O. Phishing website detection: Forest by penalizing attributes algorithm and its enhanced variations. *Arab. J. Sci. Eng.* **2020**, *45*, 10459–10470. [CrossRef]
6. Alsariera, Y.A.; Adeyemo, V.E.; Balogun, A.O.; Alazzawi, A.K. Ai meta-learners and extra-trees algorithm for the detection of phishing websites. *IEEE Access* **2020**, *8*, 142532–152542. [CrossRef]
7. Why Deep Learning over Traditional Machine Learning? Medium. 2022. Available online: https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063 (accessed on 22 June 2022).
8. Grover, R. Deep Learning-Overview, Practical Examples, Popular Algorithms | Analytics Steps. Analyticssteps.com. 2022. Available online: https://www.analyticssteps.com/blogs/deep-learning-overview-practical-examples-popular-algorithms (accessed on 22 June 2022).
9. Qazi, E.U.H.; Hussain, M.; Aboalsamh, H.; Malik, A.; Amin, H.U.; Bamatraf, S. Single Trial EEG Patterns for the Prediction of Individual Differences in Fluid Intelligence. *Front. Hum. Neurosci.* **2017**, *10*, 687. [CrossRef] [PubMed]
10. Emad-ul-Haq, Q.; Hussain, M.; Aboalsamh, H.A. Method of Classifying RAW EEG Signals. U.S. Patent 10,299,694 B1, 28 May 2019.
11. Bagui, S.; Nandi, D.; White, R.J. Machine learning and deep learning for phishing email classification using one-hot encoding. *J. Comput. Sci.* **2021**, *17*, 610–623. [CrossRef]
12. Sarker, I.H. Deep Cybersecurity: A Comprehensive Overview from Neural Network and Deep Learning Perspective. *SN Comput. Sci.* **2021**, *2*, 154. [CrossRef]
13. Feng, J.; Zou, L.; Nan, T. A Phishing Webpage Detection Method Based on Stacked Autoencoder and Correlation Coefficients. *J. Comput. Inf. Technol.* **2019**, *27*, 41–54.
14. Huang, Y.; Yang, Q.; Qin, J.; Wen, W. Phishing URL Detection via CNN and Attention-Based Hierarchical RNN. In Proceedings of the 2019 18th IEEE International Conference On Trust, Security and Privacy in Computing and Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), Rotorua, New Zealand, 5–8 August 2019; pp. 112–119. [CrossRef]
15. What Is Deep Learning and How Does It Work? SearchEnterpriseAI. 2022. Available online: https://www.techtarget.com/searchenterpriseai/definition/deep-learning-deep-neural-network (accessed on 23 June 2022).

16. Shrestha, A.; Mahmood, A. Review of Deep Learning Algorithms and Architectures. *IEEE Access* **2019**, *7*, 53040–53065. [CrossRef]
17. Do, N.Q.; Selamat, A.; Krejcar, O.; Herrera-Viedma, E.; Fujita, H. Deep Learning for Phishing Detection: Taxonomy, current challenges and Future Directions. *IEEE Access* **2022**, *10*, 36429–36463. [CrossRef]
18. Van Houdt, G.; Mosquera, C.; Nápoles, G. A review on the long short-term memory model. *Artif. Intell. Rev.* **2020**, *53*, 5929–5955. [CrossRef]
19. Su, Y. Research on Website Phishing Detection Based on LSTM RNN. In Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 12–14 June 2020; pp. 284–288. [CrossRef]
20. Aljofey, A.; Jiang, Q.; Qu, Q.; Huang, M.; Niyigena, J. An Effective Phishing Detection Model Based on Character Level Convolutional Neural Network from URL. *Electronics* **2020**, *9*, 1514. [CrossRef]
21. Singh, S.; Singh, M.P.; Pandey, R. Phishing Detection from URLs Using Deep Learning Approach. In Proceedings of the 2020 5th International Conference on Computing, Communication and Security (ICCCS), Patna, India, 14–16 October 2020; pp. 1–4. [CrossRef]
22. Yao, W.; Ding, Y.; Li, X. Deep Learning for Phishing Detection. In Proceedings of the 2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom), Melbourne, Australia, 11–13 December 2018; pp. 645–650. [CrossRef]
23. Opara, C.; Wei, B.; Chen, Y. HTMLPhish: Enabling Phishing Web Page Detection by Applying Deep Learning Techniques on HTML Analysis. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–8. [CrossRef]
24. Korkmaz, M.; Kocyigit, E.; Sahingoz, O.K.; Diri, B. Phishing Web Page Detection Using N-gram Features Extracted From URLs. In Proceedings of the 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, 11–13 June 2021; pp. 1–6. [CrossRef]
25. Tajaddodianfar, F.; Stokes, J.W.; Gururajan, A. Texception: A Character/Word-Level Deep Learning Model for Phishing URL Detection. In Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 2857–2861. [CrossRef]
26. Yerima, S.Y.; Alzaylaee, M.K. High Accuracy Phishing Detection Based on Convolutional Neural Networks. In Proceedings of the 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia, 19–21 March 2020; pp. 1–6. [CrossRef]
27. Do, N.; Selamat, A.; Krejcar, O.; Yokoi, T.; Fujita, H. Phishing Webpage Classification via Deep Learning-Based Algorithms: An Empirical Study. *Appl. Sci.* **2021**, *11*, 9210. [CrossRef]
28. Adebowale, M.; Lwin, K.; Hossain, M. Intelligent phishing detection scheme using deep learning algorithms. *J. Enterp. Inf. Manag.* **2020**. [CrossRef]
29. Zhang, L.; Zhang, P. PhishTrim: Fast and adaptive phishing detection based on deep representation learning. In Proceedings of the 2020 IEEE International Conference on Web Services (ICWS), Beijing, China, 19–23 October 2020; pp. 176–180. [CrossRef]
30. Janet, B.; Reddy, S. Anti-phishing System using LSTM and CNN. In Proceedings of the 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangaluru, India, 6–8 November 2020; pp. 1–5. [CrossRef]
31. URL 2016 | Datasets | Research | Canadian Institute for Cybersecurity | UNB. Unb.ca. 2022. Available online: https://www.unb.ca/cic/datasets/url-2016.html (accessed on 28 November 2020).
32. Mahdavifar, S.; Ghorbani, A. Application of deep learning to cybersecurity: A survey. *Neurocomputing* **2019**, *347*, 149–176. [CrossRef]
33. Chai, J.; Zeng, H.; Li, A.; Ngai, E.W.T. Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Mach. Learn. Appl.* **2021**, *6*, 100134. [CrossRef]
34. Adebowale, M.A.; Lwin, K.T.; Hossain, M.A. Deep Learning with Convolutional Neural Network and Long Short-Term Memory for Phishing Detection. In Proceedings of the 2019 13th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), Island of Ulkulhas, Maldives, 26–28 August 2019; pp. 1–8. [CrossRef]
35. Bahnsen, A.C.; Bohorquez, E.C.; Villegas, S.; Vargas, J.; González, F.A. Classifying phishing URLs using recurrent neural networks. In Proceedings of the 2017 APWG Symposium on Electronic Crime Research (eCrime), Phoenix, AZ, USA, 25–27 April 2017; pp. 1–8. [CrossRef]
36. Chen, W.; Zhang, W.; Su, Y. Phishing detection research based on LSTM recurrent neural network. In *International Conference of Pioneering Computer Scientists, Engineers and Educators*; ICPCSEE 2018: Zhengzhou, China, 2018; pp. 638–645.
37. Ariyadasa, S.; Fernando, S.; Fernando, S. Detecting phishing attacks using a combined model of LSTM and CNN. *Int. J. Adv. Appl. Sci.* **2020**, *7*, 56–67.
38. Pham, T.; Hoang, V.; Ha, T. Exploring Efficiency of Character-level Convolution Neuron Network and Long Short Term Memory on Malicious URL Detection. In Proceedings of the 2018 VII International Conference on Network, Communication and Computing–ICNCC 2018, Taipei City, Taiwan, 14–16 December 2018.
39. Lakshmi, V.; Vijaya, M. Efficient prediction of phishing websites using supervised learning algorithms. *Procedia Eng.* **2012**, *30*, 798–805. [CrossRef]

40. Malicious Url Recognition and Detection Using Attention-Based CNN-LSTM-KSII Transactions on Internet and Information Systems (TIIS) | Korea Science. Available online: https://www.koreascience.or.kr/article/JAKO201905959996575.page (accessed on 20 June 2022).
41. Zhang, Q.; Bu, Y.; Chen, B.; Zhang, S.; Lu, X. Research on phishing webpage detection technology based on CNN-BiLSTM algorithm. *J. Phys. Conf. Ser.* **2021**, *1738*, 012131. [CrossRef]
42. Jawade, J.V.; Ghosh, S.N. Phishing website detection using Fast.ai Library. In Proceedings of the 2021 International Conference on Communication information and Computing Technology (ICCICT), Mumbai, India, 25–27 June 2021. [CrossRef]
43. Tang, L.; Mahmoud, Q.H. A deep learning-based framework for phishing website detection. *IEEE Access* **2022**, *10*, 1509–1521. [CrossRef]