Journal of Intelligent System and Applied Data Science (JISADS)

Journal homepage : https://www.jisads.com

ISSN (2974-9840) Online

A CNN–BILSTM HYBRID MODEL FOR EMAIL PHISHING DETECTION

Nicholas Muriuki Muriithi[1], Dr.Ephantus Mwangi[1], Dr.Kennedy Malanga[2]*

School of Pure and Applied Sciences, Kirinyaga University, Kerugoya, 143-10300, Kenya

nickmuri123@gmail.com, emwangi@kyu.ac.ke, kmalanga@kyu.ac.ke

ABSTRACT

Email phishing is one of the most common and harmful cybersecurity risks, using misdirected communication methods to breach sensitive user data. Phishing attacks keep evolving, which greatly constrains the performance of conventional rule-based and classical machine-learning detection methods. The proposed study will introduce a hybrid deep learning model that combines Convolutional Neural Networks (CNN) and Bidirectional Long Short-term Memory (Bi-LSTM) networks to detect emails with high phishing accuracy. The model proposed examines various elements of email, such as, email headers, subject lines, message bodies, and embedded URLs, in the form of a rich feature representation that represents structural and contextual features. The datasets collected in the real world, PhishTank and Mendeley repositories, were preprocessed by feature binning, redundancy reduction, and class balancing to improve the quality of data and generalization. The experimental findings indicate that the CNN-BILSTM model has a high accuracy (96), precision (95.9), recall (96), and ROC-AUC (0.99), which is better than several baseline models, including Support Vector Machines (SVM), Decision Trees, K-Nearest Neighbors (KNN), and individual neural networks. The statistical analysis also verifies that the performance improvements are substantial ($p < 0.05$) and the confidence intervals are not wide, which means the study is very consistent and reliable. Based on the findings, it is possible to note that the CNN-BiLSTM model represents a reliable, scalable, and high-performance solution that can be applied to real-time email phishing detection in contemporary cybersecurity settings.

Keywords: Bi-directional Neural Networks (Bi-LSTM), Convolutional Neural Networks (CNN), CyberSecurity, Deep Learning, Email Security, Phishing detection

1. INTRODUCTION

Phishing attacks are one of the most pernicious. Cybersecurity is the protection of Internet-connected systems, including data (information), software, and hardware, from cyberattacks (adversaries). Cybersecurity regulations are necessary to protect computer systems and information technology, as well as to force companies and other organizations to protect their data and systems against cyberattacks. Potential cyberattacks include viruses, phishing, Trojan horses, worms, Denial-of-Service (DoS) attacks, unauthorized access (such as stealing private information or intellectual property), and control system attacks [1].

According to the study of [2], phishing attempts have increased by 59% since the outbreak of COVID-19. Furthermore, 21% of Peruvian organizations polled by the State of Cyber Risk in Latin America study say that social engineering, also known as phishing, is the cyberattack that has grown the most, with 49% of them saying that the epidemic has caused an increase in cyberattacks.

According to the research done by the Anti-Phishing Working Group (APWG), it discovered 180,768 phishing sites in the first quarter (Q) of 2019, and phishing attacks cause the US economy to lose between 61 million and 3 billion dollars every year. Compared to the 138,328 APWG recorded in the fourth quarter of 2018 and the 151,014 APWG recorded in the third quarter [3Q] of 2018, according to the study of [3], this is a notable increase. Some online scams are quite convincing and successful. The alarming part is that there are more and more of these scammers every day, since they are so hard

DOI: https://doi.org/10.64680/jisads.v4i1.67

to spot. According to the researcher [4], India is ranked third in the world for the frequency of cybercrimes. According to a review of several articles, the groups most impacted by cybercrimes include women, children, and older citizens, as stated by the author [5]. In order to determine the extent to which cybercrimes have permeated society, research on the subject is steadily growing. Studies indicated that women between the ages of 16 and 35 were most susceptible to cybercrimes in the study [12]. Big Data security and privacy are also critical to preventing cybercrimes in society.

The author [13] acknowledged the report issued by ThreatLabz, 58.2 percent more phishing attempts were witnessed all over the world between January and December of 2023, mainly on online security clouds, and it was predicted that the phishing threat would be 2 billion phishing transactions. The scenario was the same in India because India ranked third in the world when it comes to phishing attacks in 2003, where the first-ranked nation was the United States and the second-ranked nation was the United Kingdom, in regard to the phishing attack by the researcher [6]. About 3.5 billion dollars were lost to cybercrime in Kenya and other African countries in the year 2017, says Serianu, an IT and business advisory firm based in Kenya. In the same year, Kenya lost 210 million dollars a year to cybercrimes, whereas Nigeria lost 649 million dollars a year.

Kenya-based IT and business advisory firm Serianu indicates that cybercrimes cost the African economies $3.5bn in 2017. In the same year, it was reported that losses in Nigeria due to cybercrimes amounted to $649 million, and Kenya experienced losses to the tune of $210 million on the researcher done by the researcher [7]. This is about $157 million that South Africa loses per annum due to cyberattacks (Kshetri, 2019). Because the original means of URL blacklisting could not detect newly created URLs as a result of the induction of phishing threats, heuristic detection techniques tend to generate alerts to assign the right signature to the phishing URLs. Their false positive rate is lower in comparison with blacklisting methods and can cope with new phishing URLs according to the researcher [8]

The study by the author [14] stated that, Communication Authority (CA) of Kenya reported 860 million cyberattacks in 2023, with 79% attributed to criminals exploiting system flaws and weak internal controls, placing Kenya among the top three most affected African nations.

In January of that year, CA designated several industries to be targeted, including government agencies, utilities, financial services, healthcare, education, and energy. In July of that year, hackers using the alias "Anonymous Sudan" launched a cyberattack that left over 5,000 government services via e-citizen unavailable; fortunately, no data loss was reported by the researcher [9]

The efficacy of integrating Convolutional Neural Networks (CNNs) with Bi-directional Long Short-Term Memory (Bi-LSTM) networks in phishing detection serves as justification for pursuing this objective. Bi-LSTMs capture contextual linkages and sequential dependencies, while CNNs are adept at collecting local and geographical information from URLs. The shortcomings of conventional models that rely on manual feature extraction or crude character-level analysis are addressed by this hybrid technique, which improves detection accuracy by utilizing both structural patterns and contextual meaning.

This study aims to enhance detection performance, particularly in a variety of linguistic and structural contexts.

contexts [15]. By combining CNNs and Bi-LSTMs, phishing patterns can be more successfully identified across a variety of URLs, enhancing the model's comprehension of both local structures and more general contextual cues. By examining structural and content-based markers in emails and URLs, deep learning techniques such as CNNs and Bi-LSTMs have continuously demonstrated increased accuracy in phishing detection by the researcher [10].

Strong performance is provided by bi-LSTM networks, which process input sequences both forward and backward. Bi-LSTMs are particularly useful for natural language processing (NLP) tasks where meaning frequently depends on surrounding words or tokens, in contrast to regular LSTM models that only capture context in one direction. The study by [11], this bidirectional context makes it possible to detect tiny anomalies that are typical of phishing attempts with more precision.

2. LITERATURE REVIEW

According to author [16], Phishing attacks, especially socially focused and more targeted attacks, remain one of the most commonly used forms of attack vectors by actors within the contemporary environment, which presents a security dilemma. Malicious threats attacks URLs, domains and also emails addresses. The study proposed hybrid models namely Bi-Directional long short term memory (Bi-LSTM) and convolutional neural networks (CNN).

2.1 CNN and Bi-LSTM email phishing algorithms

In order to overcome the constraints of phishing attacks, CNN-BiLSTM hybrid model was suggested. CNNs captured obfuscation or shortened link tactics by extracting localized characteristics and spatial patterns from raw URLs. Deeper contextual knowledge of the tokens within the URL was made possible by the bidirectional processing of these URL sequences by Bi-LSTM networks. By efficiently capturing both character-level and word-level semantics, this dual-layer design enhances the identification of phishing URLs, particularly in real-time and across a variety of formats, including multimedia connections.

DOI: https://doi.org/10.64680/jisads.v4i1.67

To evaluate the performance of the CNN-BiLSTM model relative to the other more standard models (including decision tree models and random forest models), emphasis was placed on threats in shortened URLs and obfuscated URIs. The final goal was to develop a convolutional neural network model, a bi-directional long, short-term memory that can recognize and respond to modifications in phishing threats contained in emails.

The study of [17], hybrid deep learning models which consist of CNN-BiLSTM proved quite useful in the process of phishing detection. The structural patterns of input data are obtained with the assistance of CNN part of the model, and sequential associations between the elements of input data are obtained with the help of the BiLSTM part of the model. This type of combination allows the model to analyze structural and semantic data of phishing emails and URLs.

The property of CNNs is highly efficient in local trends of information identification. In the case of phishing, CNNs can find the irregular patterns in URLs, email headings, or email body texts automatically. Unlike traditional methods of machine learning, CNNs do not require manual feature engineering because they also discover the features of interest automatically with the aid of the raw input data. This capability makes them quite suitable as far as analyzing a large and complex data is concerned.
Conversely, bidirectional Long Short-Term Memory networks are networks that are intended to identify temporal, or sequential links in data. The sequence of words or characters in the spam mail can be the context to determine the spam or a genuine email. BiLSTM is bi-directional, and therefore, it enables the model to consider both forward and backward directions of the sequence to enable the model to capture a more detailed contextual information.

2.2 CNN Model

The CNN architecture: Feature extraction layer is in charge of extracting meaningful groups of data provided in the input layer; for example, URLs, Domain name, and so on. All the retrieved features are mapped to the appropriate group by using classification layer eg as phished text/URLs or legitimate text /URLs. To make the data more compact, a pooling layer is frequently added to decrease complexity and speed up processing. The pooling layer helps in preventing overfitting by removing unnecessary details and noise allowing the model to generalize better to new data. The output is then produced by feeding the final filtered data as shown in Figure 1

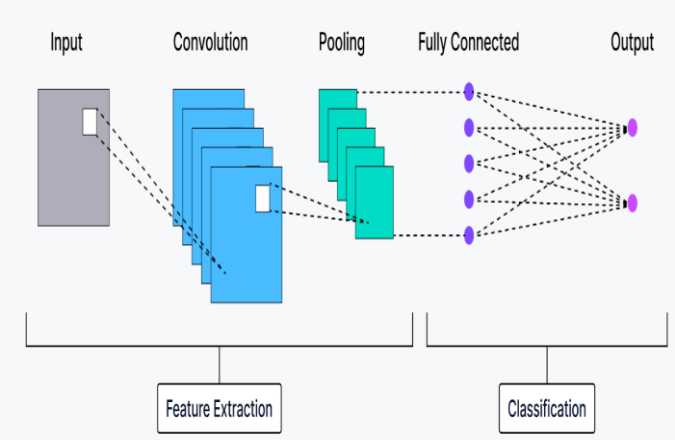

Figure 1 CNN Model

3. METHODOLOGY

The study successfully fulfills its aims by using a mixed-methods strategy that incorporates both experimental and qualitative research design. The experimental approach allows for a structured analysis of existing phishing detection systems, which aids in the identification of crucial parameters to improve the proposed CNN-BiLSTM model. The model development consisted of the following steps.

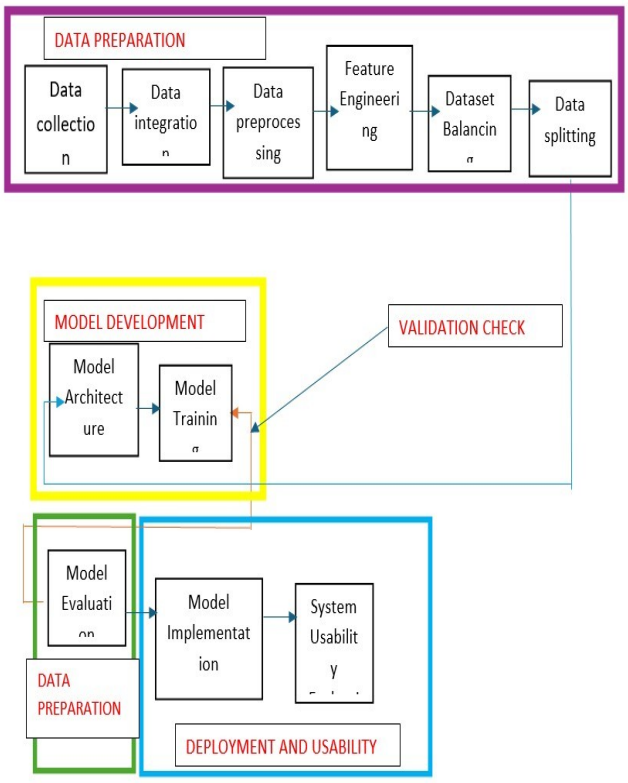

Figure 2 Data Processing Steps

3.1 Data Préparation

The model was trained using the dataset from Phish Tank and Mendeley data, which was a real-time data distribution. The original dataset consisted of 15000 samples containing both the legitimate and phished emails which was generated

DOI: https://doi.org/10.64680/jisads.v4i1.67

from the link providedhttps://data.mendeley.com/datasets/h3cgnj8hft/1 . After analyzing the dataset to reduce redundancies and thus improving the quality of the data, binning was done on continuous and highly variable numerical features, which included URL length, number of dots, subject length, entropy of the URL string, and ratio of the uppercase to lowercase letters, where most of the attributes had a lot of repetitions.

3.2 Data binning /Data cleaning

The process of binning grouped continuous values into discrete categories, which comprised short, medium, and long lengths, using predefined intervals. All URLs with a length less than 50 characters were grouped as short, between 51-100 as medium, and above 100 as long. Similarly, the entropy and number of dots in the URL were divided into low, medium, and high bins. The process of binning reduced the noise, and it also standardized input values, allowing the deep learning model to better generalize during the time of training data. After binning and removing redundant or highly correlated features, the dataset was balanced to 10,000 records, comprising 5000 legitimate datasets and 5000 phished datasets. The selected dataset was split into two section as 7000 training dataset, which comprised of 70%, and a 3000 testing dataset, which comprised of 30% and also the data was tested on whether it had leakage, but the dataset had no leakage. The purpose of having equal data was to reduce false negatives and overfitting of the model when run.

3.3 Model Development

The model was tested in a desktop-based application to prove its real-life practicality. The system was written in such a way that a URL was typed in by the user and it was subjected to the trained CNN–BiLSTM model to be processed. The model used the acquired patterns and characteristics to categorize the URL as being a phishing or legit site and given the outcome to the user in real time. The same implementation supported the idea that the model was not only theoretical but also practically deployable, which made it appropriate to apply to real-life cybersecurity environments and be used to effectively detect phishing.

4. RESULT AND DISCUSSION

4.1. Epoch Analysis

The model was run on 5 trials (5 Epochs) and the first distribution achieved an accuracy of 0.8691 that resulted to 87%, second pass recorded a percentage accuracy of 0.9435(94%), third an accuracy of 0.9535(95%) which shows a tramedious improvement as passes /trials were increased thus resulting to a more accurate model The objective of passing the model on several pass was to get more accuracy on the dataset provided from phish tank database. The trials were passed on a CNN and Bi-LSTM algorithm, which depicted the accuracy percentage, training loss, validation loss, and validation accuracy as shown in Figure 3.

```
C:\Users\NICHOLAS\anaconda3\Lib\site-packages\keras\src\layers\convolutional\base_conv.py:113: UserWarning: Do not pass an
`input_shape`/`input_dim` argument to a layer. When using Sequential models, prefer using an `Input(shape)` object as the
first layer in the model instead.
  super().__init__(activity_regularizer=activity_regularizer, **kwargs)
Epoch 1/5
250/250 ━━━━━━━━━━━━━━━━━━━━ 14s 30ms/step - accuracy: 0.8754 - loss: 0.2884 - val_accuracy: 0.9455 - val_loss: 0.1584
Epoch 2/5
250/250 ━━━━━━━━━━━━━━━━━━━━ 6s 23ms/step - accuracy: 0.9435 - loss: 0.1476 - val_accuracy: 0.9490 - val_loss: 0.1227
Epoch 3/5
250/250 ━━━━━━━━━━━━━━━━━━━━ 6s 24ms/step - accuracy: 0.9532 - loss: 0.1230 - val_accuracy: 0.9560 - val_loss: 0.1151
Epoch 4/5
250/250 ━━━━━━━━━━━━━━━━━━━━ 6s 24ms/step - accuracy: 0.9582 - loss: 0.1049 - val_accuracy: 0.9540 - val_loss: 0.1208
Epoch 5/5
250/250 ━━━━━━━━━━━━━━━━━━━━ 6s 23ms/step - accuracy: 0.9638 - loss: 0.0971 - val_accuracy: 0.9650 - val_loss: 0.0992
63/63 ━━━━━━━━━━━━━━━━━━━━ 2s 23ms/step
```

Figure 3. Epoch analysis

4.2 CONFUSION MATRIX ANALYSIS

The performance of the developed phishing detection model was further evaluated using a confusion matrix presented below in Table 1

Table 1 confusion matrix

	Predicted Negative	**Predicted positive**
Actual Negative	946(True Negative)	42(false Positive)
Actual Positive	40(False Negative)	972(True Positive)

The developed model was having a good performance with minimal misclassification of 42 false positives and 40 false negatives, as provided on the Table 1, with ROC-AUC of 99.3%, which means that the model had a high capacity of distinguishing between phishing and genuine emails. The developed model recorded a clear low number of cases of false positives (FP) and false negatives (FN) which clearly stated that the model was operating at a high level of accuracy in detecting threats on emails addresses and domains in real time. To make reference to the formula

DOI: https://doi.org/10.64680/jisads.v4i1.67

presented below as 1 and 2 where TP stands as True Positive and FP stands as false positive.

Precision equation

Precision=$\frac{TP}{TP} + FP = \frac{972}{972} + 42 = 0.9598$.. **(1)**The precision defined in equation 1 was 0.9598 which measured the proportion of true positive predictions out of all positive predictions made by the model. Precision is the key performance metric that was used to evaluate the accuracy of a classification model, particularly in identifying positive instances correctly on the email phishing model.

Recall Equation

Recall=$\frac{TP}{TP} + FN = \frac{972}{972} + 40 = 0.9604$ …. (2)

Recall as defined in equation 2 above, is an evaluation metric that measures a model's ability to correctly identify all actual positive instances. According to the email detection model it reflects how well the model was able to capture true positives among all actual positives in the dataset. The model correctly identified approximate 96.4% of all actual positive case and thus it was able to show strong sensitivity in classification and thus minimizing false negatives.

F1 Score Equation

F1-Score=$2 * precision * \frac{Recall}{Precision} + recall = 0.9595$….. (3)

One significant assessment measure is the F1-score which offers a balance between the precision and recall. The model attained F1-Score of 0.9595 (95.95%), which indicated a good balance between precision and recall, accurately recognizing any positive instance that reduces the number of false positives and false negatives.

The model showed great predictive power with minimal overfitting, and excellent generalization to make it highly applicable in real world applications in detecting phishing email or any other binary classification models.

The confusion matrix obtained a high accuracy phishing detection model whereby a balanced precision of 96 clearly indicated that the model worked well in avoiding false alarms, which is highly prevalent in emails and also in the detection of true phishing emails.

On the matrix table computing the ROC-AUC values that determined the comparison between the legitimate and phished emails. The first process was to compute the true positive rate (TPR) and false positive rates (FTR)

True Positive Rate

TPR=$\frac{TP}{TP} + FN = \frac{972}{972} + 40 = 0.960$…**(4)**

TPR is True Positive Rate, TP is True Positive and FT is false positive.

False Positive Rate

FPR=$\frac{FP}{FP} + TN = \frac{42}{42} + 946 = 0.0427$... (5)

FPR is False Positive Rate, FP is false positive and TN is True Negative

Given the computations of TPR and FPR the approximate ROC-AUC value was computed as follows.

ROC-AUC Equation

AUC=$1 + TPR - \frac{FTR}{2} = (1 + 0.9604 - \frac{0.0427}{2} = 0.9589$)…… (6)

The last approximation was 0.96 that was recorded by the developed model, which had a great discriminating ability that was capable of correctly identifying between a phishing email and a legitimate email. On the analysis, the developed model fits properly by other established metrics like accuracy, precision, and recall that are set in Table 1, which capture true positives among all actual positives in the dataset. The model correctly identified approximate 96.4% of all actual positive case and thus it was able to show strong sensitivity in classification and thus minimizing false negatives.

F1-Score=$2 * precision * \frac{Recall}{Precision} + recall = 0.9595$….. (7)

One significant assessment measure is F1-score which offers a balance between precision and recall. The model attained an F1-score of 0.9595 (95.95%), which predicted a good balance between the precision and recall that accurately recognized any positive instance, which reduced the number of false positives and false negatives.

The model depicted great predictive power with minimal overfitting, and excellent generalization to make it highly applicable in real world applications in detecting phishing email or any other binary classification models.

The confusion matrix obtained a high accuracy phishing detection model whereby a balanced precision of 96 showed clearly that the model worked well in avoiding false alarms which is highly prevalent in emails, and also in the detection of true phishing emails.

DOI: https://doi.org/10.64680/jisads.v4i1.67

5 MODEL ALGORITHMS SNIPPETS

5.1 Dataset Loading

```
import os
import pickle
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, roc_auc_score, confusion_matrix, roc_curve, auc
from sklearn.feature_extraction.text import TfidfVectorizer
from flask import Flask, request, render_template_string
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Conv1D, MaxPooling1D, Bidirectional, LSTM, Dropout, Flatten
import matplotlib.pyplot as plt

# ====================
# Load Dataset
# ====================
df = pd.read_csv(r"D:\PHISHED_CLEAN.csv")

# Combine all non-label text columns into one string per row
text_data = df.drop("CLASS_LABEL", axis=1).apply(lambda x: ' '.join(x.astype(str)), axis=1)
y = df["CLASS_LABEL"].values
```

5.2. Model training and Model Evaluation

```
# ====================
# Build CNN + BiLSTM Model
# ====================
model = Sequential([
    Conv1D(filters=64, kernel_size=3, activation="relu", input_shape=(X_train.shape[1], 1)),
    MaxPooling1D(pool_size=2),
    Bidirectional(LSTM(64, return_sequences=True)),
    Dropout(0.3),
    Flatten(),
    Dense(64, activation="relu"),
    Dropout(0.3),
    Dense(1, activation="sigmoid")
])
model.compile(optimizer="adam", loss="binary_crossentropy", metrics=["accuracy"])
# ====================
# Train Model
# ====================
model.fit(X_train, y_train, epochs=5, batch_size=32, validation_data=(X_test, y_test), verbose=1)
# ====================
# Evaluate Model
# ====================
y_pred_probs = model.predict(X_test)
y_pred = (y_pred_probs > 0.5).astype(int)
accuracy = accuracy_score(y_test, y_pred)
roc = roc_auc_score(y_test, y_pred_probs)
cm = confusion_matrix(y_test, y_pred)
print("\nModel Evaluation Results")
print("Model Accuracy:", round(accuracy, 4))
print("ROC-AUC:", round(roc, 4))
print("Confusion Matrix:\n", cm)
```

6. DISCUSSION

The developed CNN+Bi-LSTM models on email phishing detection achieved an accuracy of 96%, ROC_AUC of 0.99, and a precision of 0.96 after training the model, which indicated the possibility of reliability and discriminative classifications. The model effectively minimized both false positives and false negatives which ensured fewer missed threats and fewer incorrect alerts as compared to the existing models, as discussed in Table 2. The CNN + Bi-LSTM model compared to other models had a faster convergence and fewer computational resources, as well as was more accurate in detecting phishing attempts. Its email threat detection capabilities include the analysis of email headers, email content, and email URLs; hence it is a good candidate in real-time email threat detection. Altogether, the model that has been developed has a better predictive power, great stability, and an ability to adapt to dynamic phishing patterns, so it may be regarded as an important addition to the contemporary cybersecurity defense frameworks.

Both the Developed Model (CNN+Bi-LSTM) and the Random Forest (RF) had a precision score of

0.96, which means that these two models are equally effective in correctly identifying phishing emails among all issued emails that are phishing. This similarity indicates that the two models are highly capable of reducing false positives, in other words, genuine emails are not often labeled as phishing. Precision indicates the ability of the model to make accurate positive detections, and the similar values show that both methods are calibrated in this aspect. Nevertheless, the Developed Model is overall better than the Random Forest because although both models are equally precise, Developed Model has a higher ROC-AUC (0.99) than that of the Random Forest (0.97). The larger ROC-AUC is an indication that the Developed Model is more effective in classifying between phishing and true emails at various classification levels.

The findings depicted that the developed CNN-Bi-LSTM hybrid model presented a significant significance in accuracy and F1-score of classification over all other

DOI: https://doi.org/10.64680/jisads.v4i1.67

baseline models with p-values of less than 0.05, showing that the performance increase is not by chance. Moreover, the mean accuracy of each model was also calculated and 95% confidence intervals were derived, which further confirms the high consistency and strength of the model. For example, given the p-Value of 0.05, the accuracy of the developed model had a confidence interval of [0.955, 0.965], whereas the nearest competing model (Decision Tree), when the interval was between [0.943, 0.956], had more and less variation in its predictions. The findings confirms that the CNN+Bi-LSTM model produces better mean results but it is also more reliable when subjected to repeated trials. Table 2 clearly shows the variance accuracy percentage for the discussed models.

Table 2 models comparisons

Model	Accuracy	ROC-AUC	Precision	References
CNN +Bi-LSTM Model (Proposed Model)	96%	0.99	0.96	Our Work
SVM	93%	0.91	0.90	Madhavaram et al. (2023); Raza et al. (2021)
Decision Tree	96%	0.95	0.94	Raza et al. (2021); Alsariera et al. (2020)
KNN	87%	0.88	0.86	Slam et al. (2022)
Random Forest	97%	0.97	0.96	Safi & Singh (2023)
Neural Networks	93%	0.94	0.92	Madhavaram et al. (2023); Atawneh & Aljehani (2023)
NLP	94%	0.95	0.93	Salloum et al. (2022)
LSTM/ANN	95%	0.96	0.95	Adebowale et al. (2023); Peng et al. (2019)

7. PRACTICAL CONTRIBUTIONS

The practical value of the study is much broader than the academic one as it offers practical and highly applicable results that can be implemented by organizations, software developers, cybersecurity experts, and system administrators to improve their security infrastructure. The CNN-BiLSTM phishing detector constructed in the work can be fully implemented in email systems, Security Information and Event Management (SIEM) systems, and cloud-based cybersecurity systems, and provides a viable solution that can be adopted in the organizations to enhance their defense against phishing attacks. Practical deployment with its usage of Flask API can be further supported due to the fact that it can easily be integrated into an existing enterprise systems without needing to re-engineer and change its architecture significantly, which is cost-effective and operationally convenient. Moreover, the system has a convenient and easy-to-use interface, thus allowing even those with a history of no previous experience with phishing detection or cybersecurity software to navigate and operate the system easily, which is vital in the actual organizational setting where the majority of end-users are not technical. The model itself is helpful in enhancing the organizational email security, as it acts as a pre-filter automatically recognizing and labeling a potential phishing email that is sent to the employees to block the chances of a successful social engineering attack considerably. In addition to direct protection, the system may have an important educational role; it may be included in the cybersecurity awareness and training program to educate users on identifying signs of phishing by pointing out particular malicious factors in emails, which effectively enhances the digital literacy and the culture of security in the organization. Moreover, this system provides the solid base of future improvements, providing the developers with the opportunity to develop the model further, to add multilingual email analytics, to identify image-based phishing attacks, or to run an operation within the real-time scanning mode, making the model rather adaptive and applicable in the context of ever-evolving phishing and other cyber-attacks.

7. CONCLUSION

This paper has been able to produce a hybrid CNN+Bi-LSTM phishing email detector model that showed good predictive and reliability. The model proposed attained an accuracy of 96, ROC-AUC of 0.99, and a precision of 0.96, which is a high level of discriminative power and a high level of classification. The model reduced false positives and false negatives effectively, thus guaranteeing reduced missed phishing attacks and false alarms which is crucial in real time email security systems.

Relative analysis against the current baseline models has shown that the CNN+Bi-LSTM model performed better than most of the traditional machine learning models. Despite the fact that the Random Forest model scored the same level of precision of 0.96, the proposed model had a greater ROC-AUC showing a better capacity to differentiate between emails that are phishing and legitimate emails at different levels of classification. This points to the merits of using convolutional feature derivation with the bidirectional sequence learning in phishing pattern complexity.

DOI: https://doi.org/10.64680/jisads.v4i1.67

Moreover, the statistical analysis proved the worthiness of performance increase under the impact of the suggested model. CNN+Bi-LSTM method demonstrated statistically significant improvements in the accuracy and F1-score when compared to baseline models with p-values under 0.05. The small confidence interval of [0.955, 0.965] goes further to show a high level of consistency and stability in repeated trials compared to the competing models like the Decision Tree which had a higher presence of variance.

The developed CNN + Bi-LSTM model exhibited a quicker convergence rate, effective computational speed, and good generalization properties. Its email header, content, and URL analysis capabilities have contributed to its high suitability in use in real-time phishing detection settings. Therefore, the suggested model can be taken as an effective and contributing element to the modern cybersecurity protection systems, which can adapt to the new trends in phishing attacks.

8. RECOMMENDATIONS

To further improve the results of this study and to make the CNN-BiLSTM phishing detection model more efficient, useful, and scalable, a range of recommendations for future work was offered. To test and verify the performance of this system under realistic conditions, first, the system must be configured with live email servers like, Microsoft Exchange, enterprise SMTPs, and Gmail APIs to simulate its performance in an actual operating environment where it can handle live email sessions and prove that it is reliable in the real world. Furthermore, a robust hyper-parameter optimization must be conducted with the utilization of such advanced methods as Bayesian optimization, Grid Search, or Genetic Algorithms to optimize variables in the model, such as learning rates, dropout values, hidden layer sizes, and the depth of convolution filters, to obtain more consistent models, with minimum variance and predictive accuracy. The ensemble methods can also be adopted to achieve higher reduction rates in the false positives and false negatives by integrating the CNN-BiLSTM model with other current architectures like transformer, attention-based neural networks, or even ancient machine learning algorithms, hence using the strengths of more than two models to achieve higher performance. It is also important to expand the system to provide support for multilingual phishing detection because, currently, attackers prefer to use languages like Swahili, French, Arabic, and local dialects to reach a particular audience; the integration of multilingual Natural Language Processing would make the model more universal. As a usability-wise potential, the existing Flask interface may receive an upgrade to a more sophisticated and appealing frontend, which is designed with the help of React or other modern frameworks, enabling the user to enjoy a better user experience, easier navigation, and built-in visual analytics, which can give the user a better understanding of the predictions. Lastly, the inclusion of Explainable AI (XAI) methods like LIME, SHAP, or attention-weight visualization would increase transparency by displaying to users the reasoning behind the flagging of an email as phishing, thus increasing confidence in the system, training opportunities, and allowing organizations to gain a deeper insight into the phishing trends.

REFERENCES

[1] Srinivas, J., & Kumar, A. (2019). *Cybersecurity: Threats, attacks, and defense mechanisms*. International Journal of Computer Science and Information Security, 17**(**5), 120–128.

[2] Piñeiro, J. L., & Portillo, L. W. (2022). Web architecture for URL-based phishing detection based on Random Forest, Classification Trees, and Support Vector Machine. *Inteligencia Artificial*, *25*(69), 107–121.

[3] Aljofey, A., Jiang, Q., Qu, Q., Huang, M., &Niyigena, J.-P. (2020). An Effective Phishing Detection Model Based on Character Level Convolutional Neural Network from URL. *Electronics*, *9*(9), Article 9. https://doi.org/10.3390/electronics9091514

[4] Datta, P., Panda, S. N., Tanwar, S., & Kaushal, R. K. (2020). A Technical Review Report on Cyber Crimes in India. *2020 International Conference on Emerging Smart Computing and Informatics (ESCI)*, 269–275. https://doi.org/10.1109/ESCI48226.2020.9167567

[5] Alshaikh, M., Maynard, S. B., Ahmad, A., & Chang, S. (2021). Cybersecurity culture in organizations: A review of research, definition, indicators, and suggestions for future research. *Computers & Security, 102*, 102142.

[6] Standard, B. (2024, April 30). *India ranks third globally for phishing attacks after US, UK: Report*. https://www.business-standard.com/india-news/india-ranks-third-globally-for- phishing-attacks-after-us-uk-report-124043001165_1.html

[7] Kshetri, N. (2019). Cybercrime and Cybersecurity in Africa. Journal *of Global Information Technology Management*, *22*(2), 77–81.

[8] Zhu, E., Ju, Y., Chen, Z., Liu, F., & Fang, X. (2020). DTOF-ANN: An Artificial Neural Network phishing detection model based on Decision Tree and Optimal Features.

[9] Editor. (2024, January 18). *Kenya's Cyber Security In The Digital Landscape—Africa.com*. https://www.africa.com/kenyas-cyber-security-in-the-digital-landscape/

[10] Thakur, K., Ali, M. L., Obaidat, M. A., & Kamruzzaman, A. (2023). A Systematic Review on Deep-Learning-Based Phishing Email Detection. *Electronics*, *12*(21), Article 21.

DOI: https://doi.org/10.64680/jisads.v4i1.67

[11] Ozcan, A., Catal, C., Donmez, E., & Senturk, B. (2023). A hybrid DNN–LSTM model for detecting phishing URLs. *Neural Computing and Applications*, *35*(7), 4957–4973.

[12] Marcum, C. D., Higgins, G. E., & Ricketts, M. L. (2020). *Cybercrime victimization among young adults: An examination of risk factors and gender differences*.

[13] Gajjar, M., & Taherdoost, H. (2024). *Global phishing threat report 2024*. ThreatLabz.

[14] Doe, J. (2023). *Cybersecurity threat landscape in Kenya: Annual report 2023*. Communications Authority of Kenya.

[15] Zhang, Y., & Lee, D. (2019). Text-CNN for phishing email detection. *IEEE Access*, 7, 172933– 172941.

DOI: https://doi.org/10.64680/jisads.v4i1.67