International Journal of Informatics Engineering and Computing (IJIMATIC)
Vol. 3, No. 1, 2026
E-ISSN: 3090-9112

Stepping up Phishing Detection using Convolutional Neural Network Algorithm

Beti Fatima Markis[1], Safaet Hossain[2]

Abstract
Phishing websites continue to pose a significant cybersecurity threat by deceiving users into disclosing sensitive information, making accurate and reliable detection methods increasingly important. This paper evaluates the effectiveness of several machine learning algorithms for phishing website detection, including Convolutional Neural Network (CNN), Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbor (KNN), Gaussian Naïve Bayes (GNB), and Gradient Boosting (GBoost). The proposed approach first analyzes the CNN learning process using training and validation accuracy and loss curves to assess model convergence and generalization capability. Subsequently, the performance of all classifiers is evaluated using confusion matrices, accuracy, precision, recall, and F1-score. Experimental results show that the CNN achieves the highest overall accuracy of 93%, with balanced precision, recall, and F1-score of 0.93 for both legitimate and phishing website classes. SVM, KNN, and GBoost also demonstrate competitive performance with 91% accuracy, whereas Decision Tree achieves 89% accuracy. In contrast, Gaussian Naïve Bayes produces the weakest performance with an accuracy of only 51%, indicating that its feature independence assumption is inadequate for modeling the complex characteristics of phishing websites. These findings demonstrate that deep learning, particularly CNN, provides more accurate and balanced phishing detection than conventional machine learning methods. The proposed approach offers an effective solution for practical phishing detection and has the potential to strengthen cybersecurity systems by reducing both false positives and false negatives. Future work will investigate advanced deep learning architectures, explainable artificial intelligence techniques, and larger real-world phishing datasets to further improve detection accuracy, robustness, and model interpretability.

Keywords:
Phishing Detection, Cybersecurity, CNN, Deep Learning

1. Introduction

Phishing detection was chosen as the primary subject of the study due to the increasing prevalence of phishing scams and the need for rapid and low-cost prevention measures[1]. CNN played a role in the study by being used as a deep learning-based technique for identifying phishing websites. Multiple variants of CNN, including one-layer CNN with various-sized kernels, were deployed in parallel to extract multi-level features [2]. CNN was also combined with other techniques such as long short-term memory (LSTM) and self-attention to improve the accuracy of phishing website detection[3]. The use of CNN in the study allowed for the construction of classifiers that achieved high accuracy rates, ranging from 95.6% to 99.2% [4] .

Phishing attacks can lead to financial losses and loss of personal data for internet users. When successful, these attacks can cause psychological, emotional, and financial damage to the victims, with long-term impacts [5]. Phishing is a form of identity theft where attackers deceive users into revealing sensitive information such as login credentials and credit card details[6]. The attackers use techniques like spoofed emails and fake websites to obtain personal information. Users are tricked into entering their confidential information into these

Corresponding Author: Beti Fatima Markis (bety.fatima@gmail.com)
1 Beti Fatima Markis, Department of Informatics, Universitas Respati Yogyakarta, Indonesia
2 Safaet Hossain, University Tenaga Nasional, Malaysia

phishing websites, allowing the attackers to steal their credentials[7]. Phishing attacks can result in the theft of personal or confidential information, leading to financial losses and potential identity theft [8]. These attacks can also cause brand reputation damage and loss of customer confidence in e-commerce and online banking.

Different machine learning techniques can produce different detection patterns in terms of false positives and false negatives. For example, in the context of rice blast disease detection, a review paper found that various machine learning algorithms, such as Naive Bayes, LSTM RNN, Random Forest Classifiers, Support Vector Machines, K Means, Decision Tree, and Convolutional Neural Networks, were evaluated. However, the future scope of improvement was identified for algorithms like Naive Bayes and Recurrent Neural Networks [9]. In the context of AI-generated text detection, statistical experiments showed that false positive and false negative detection of generative AI varied depending on the text source. Literature parts had higher likelihoods of falsely demonstrating AI-generated text, while genuine texts and AI-generated texts had noticeable overlap in their distributions, leading to the possibility of type I and type II errors [10]. In the context of ransomware detection, different machine learning models, such as Artificial Neural Network (ANN) and Random Forest, showed varying performance in terms of precision and false negatives [11]. Therefore, it can be concluded that different machine learning techniques can indeed produce different detection patterns in terms of false positives and false negatives.

The use of specific features in datasets has a significant impact on the performance of machine learning techniques in phishing detection. Feature selection techniques based on the URL have been applied to improve the detection process [12]. The Polymorphic property of features, where the same feature provides different interpretations when considered in different parts of the URL, has been utilized to enhance the accuracy of phishing URL detection. Machine learning classifiers operating on URL features have been used to detect phishing URLs efficiently [13].Tuning factors such as data balancing, hyper-parameter optimization, and feature selection have been shown to improve the accuracy of machine learning algorithms in phishing detection [14]. Various machine learning techniques, including Extreme Gradient Boosting, Decision Tree, Logistic Regression, Random Forest, and Support Vector Machine, have been used to identify phishing websites, with Random Forest achieving high accuracy rates.

Machine learning techniques, such as Random Forest (RF) and XGBoost, have been used to detect phishing attacks with high accuracy[15] [16]. Feature selection plays a crucial role in improving the performance of these models [17]. Factors such as the number of instances and features in the dataset can affect the performance of machine learning classifiers. Additionally, the nature of the features used, including content-based, URL lexical-based, and domain-based features, can influence the model's ability to identify phishing websites. CNN performance in detecting phishing attacks on this type of dataset is not explicitly mentioned in the provided abstracts. However, it is worth noting that CNNs have been widely used in various domains, including cybersecurity, and can potentially be applied to detect phishing attacks as well.

Phishing detection on websites and securing internet users can be improved through various experiments conducted in the research papers. One approach is to create machine learning models using the phishing website prediction dataset, which aims to improve performance by enhancing features on different models. Another method involves utilizing fuzzy, neural network, and data mining techniques to effectively detect phishing sites [18]. Additionally, the use of URLs as a dataset and extracting features from them can help determine if a website is phishing or not, with the Multilayer perceptron algorithm achieving high accuracy in this regard. Furthermore, a methodology that combines URL, HTML, and web technology features can be employed for phishing website detection, along with the development of a comprehensive dataset for benchmarking [19]. Lastly, a three-phase

attack detection system based on web traffic, web content, and URL classification, using a recurrent neural network, can accurately detect phishing incidents [20].

CNN performance is superior in detecting phishing attacks on the dataset compared to traditional machine learning techniques [21]. The proposed CNN-based system achieved an accuracy of 99.2% in detecting phishing websites [22]. In contrast, traditional machine learning techniques such as Random Forest, Decision Tree, and Logistic Regression achieved accuracies ranging from 93.66% to 98.80% [23] [24]. The CNN-based approach outperformed these techniques in terms of accuracy [25]. Additionally, the CNN-based system was found to be the most accurate model among the four models evaluated, including ANNs, SVM, and decision trees (DTs) . Therefore, based on the available information, it can be concluded that CNN performance is superior in detecting phishing attacks on this dataset compared to traditional machine learning techniques.

2. Related Works

The related literature documents the latest developments in phishing detection on websites through various approaches such as machine learning, deep learning, and hybrid models. Machine learning techniques have been widely applied, with the Random Forest Classifier being commonly used [26]. Deep learning models, specifically CNN, have achieved high accuracy in detecting phishing websites [27]. Researchers have also developed datasets specifically for phishing detection, such as the Phishing Index Login Websites Dataset (PILWD), which includes a wide variety of data for testing and comparing approaches [28]. Additionally, web technology features have been utilized to improve phishing detection accuracy [29]. Support Vector Machine (SVM) is another machine learning technique used for phishing website detection, which has shown promising results in terms of accuracy [30]. Overall, the literature highlights the advancements in machine learning, deep learning, and hybrid models for effective phishing detection on websites.

Machine learning algorithms have been widely used in the context of phishing detection. These algorithms analyze various features of websites, such as URL structure, website content, and the presence of specific keywords or patterns, to determine the likelihood of a website being a phishing site [31] [32]. The use of machine learning classifiers, such as SVM, has been effective in detecting phishing URLs based on URL features alone [33]. Some solutions focus on URL-based algorithms, while others consider website content [34]. The proposed approach integrates techniques like blacklist interception, whitelist filtering, and machine learning prediction to improve accuracy and reduce false alarm rates [35]. Different machine learning techniques, including XGBoost, DT, LR, RF, and SVM, have been used to identify phishing websites, with RF achieving high accuracy rates . Overall, the use of machine learning algorithms in phishing detection has progressed over time, with advancements in feature extraction, model selection, and performance evaluation.

CNN have been applied in research related to phishing detection on websites. Several papers propose deep learning approaches using CNN for this purpose. One paper compares 1D CNN and 2D CNN with three feature types (URL-based, content-based, and third-party services-based) and finds that 1D CNN is more adequate for phishing detection, achieving a high accuracy of 96.76% . Another paper proposes three deep learning-based techniques, including CNN, LSTM-CNN, and LSTM, and demonstrates their accuracy as 99.2%, 97.6%, and 96.8% respectively, with the CNN-based system being superior [36]. These studies show that CNN algorithms can effectively detect phishing websites, achieving high accuracy rates in the range of 96.76% to 99.2%.

CNN algorithms offer several advantages in phishing detection compared to other methods. Firstly, CNN-based strategies have been shown to achieve improved detection and accuracy in identifying phishing attempts . This is because CNN models can effectively analyze and extract relevant features from phishing emails, leading to more reliable

detection. Additionally, CNN algorithms can handle the complex and dynamic nature of phishing attacks by capturing patterns and variations in email content. Furthermore, CNN-based approaches can efficiently process large datasets and extract important features, making them suitable for real-time phishing detection [37]. The experimental findings also demonstrate the high accuracy of CNN-based systems in detecting phishing websites [38]. Overall, the use of CNN algorithms in phishing detection offers improved accuracy, efficient processing, and the ability to handle the dynamic nature of phishing attacks.

Machine learning techniques, such as Decision Tree, Random Forest, Support Vector Machine, Naïve Bayes, and XGBoost, have been extensively used in research related to phishing detection [39]. These techniques analyze various features of websites, including URL structure, website content, and the presence of specific keywords or patterns, to identify phishing sites. Feature selection techniques based on the URL have been applied to improve the detection process, resulting in improved accuracy [40]. The use of machine learning algorithms, such as Random Forest and XGBoost, has shown promising results in accurately detecting phishing websites, with accuracy rates ranging from 86.8% to 98.80%. These techniques can be utilized in modern web browsers to enhance security and protect users from falling victim to phishing attacks.

There are two studies that compare the performance of different machine learning techniques in phishing detection. A study by Gupta et al. focuses on detecting phishing attacks using lexical features and machine learning classifiers such as Random Forest, achieving a highest accuracy of 99.57% [41]. Another study by Alsufyani et al. uses natural language processing and machine learning algorithms including K-Nearest Neighbors, Decision Tree, and AdaBoost to detect text phishing, with three models obtaining considerable values. Both studies evaluate the performance of these techniques based on accuracy and other metrics.

The use of special features in datasets can significantly affect the performance of phishing detection techniques. Feature selection methods, such as Information Gain algorithm and wrapper feature selection techniques, are employed to reduce the number of features and improve classification accuracy [42] [43] [44]. In the case of phishing detection, the selection of discriminative features plays a crucial role in identifying phishing websites accurately. Different combinations of features, including those gathered from external services and web page contents, are evaluated to determine their effectiveness in detecting phishing attempts [45]. The performance of classifiers, such as NaiveBayes, ANN, DecisionStump, KNN, J48, RandomForest, support vector machine (SVM), and multi perceptron (MLP), is assessed based on the selected features . The results show that the choice of features and feature selection methods can significantly impact the accuracy of phishing detection techniques, with some classifiers outperforming others .

There have been previous attempts at integrating CNN algorithms with traditional machine learning techniques to improve phishing detection. Manikas et al. proposed a method that enhances steganographic approaches by including a steganographic assistant convolutional neural network (SA-CNN). Chinnasamy et al. also used machine learning models, such as Random Forest and SVM, to classify whether a link is phishing or non-phishing based on input features like web traffic and URL. Shirazi et al. proposed two GAN-based approaches, AAE and WGAN, to synthesize phishing and legitimate samples for training classifiers with higher performance and resistance to adversarial attacks. Uplenchwar et al. developed a phishing attack detection system for text messages using ML techniques like Naive Bayes, Support Vector Classification, Random Forest, and KNN [46]. These studies demonstrate the integration of CNN algorithms with traditional machine learning techniques to improve phishing detection.

Machine learning techniques in phishing detection address challenges such as false positives and false negatives by using features extracted from phishing and legitimate emails to train classifiers. These classifiers are designed to detect phishing URLs based

on URL features only[47]. Feature selection techniques based on the URL are applied to improve the detection process, ranking and selecting the most relevant features [48]. Different machine learning classifiers, such as Naïve Bayes, Random Forest, and XGBoost, are used to identify phishing websites based on the selected features. The accuracy rates of each algorithm are compared to determine the most effective machine learning technique [49]. The proposed methodology achieves high accuracy rates, with Random Forest providing an accuracy of 98.80% on one dataset. These machine learning techniques help in accurately detecting phishing attacks and reducing false positives and false negatives.

CNN algorithms have been widely used in the context of phishing detection on websites. However, there are some challenges and constraints encountered in their use. One challenge is the computational complexity of region-based and regression-based target identification methods, which extract picture characteristics from the CNN backbone and apply sliding window procedures to candidate areas [50]. Another challenge is the biases issue in LSTM, which treats later features more important than the former ones . Additionally, the dynamic nature of machine learning techniques used for phishing detection requires careful evaluation and comparison of different models [51]. Despite these challenges, researchers have proposed various deep learning-based techniques, such as 1D CNN and LSTM-CNN, which have shown high accuracy in detecting phishing websites [52] .

Previous research has influenced the development of phishing detection techniques, particularly in terms of machine learning algorithms. Researchers have used machine learning technology, such as GBoost, Decision Tree (DT), Logistic Regression (LR), Random Forest (RF), and Support Vector Machine (SVM), to identify phishing websites [53]. They have also designed systems to extract features from phishing emails and URLs, using SVM classifiers to detect phishing websites based on URL features [54]. Incorporating different machine learning and deep learning algorithms, including SVM, Gradient Boosting Machine (GBM), and random forest, researchers have presented techniques for identifying phishing websites [55]. Structural properties of phishing website URLs have been analyzed, and machine learning algorithms have been trained to identify unknown URLs [56]. These research findings have contributed to the development of accurate and efficient phishing detection techniques using machine learning algorithms [57] .

Recent trends and innovations in phishing detection include the use of machine learning and big data technologies for comprehensive analysis and detection of phishing attacks [58]. Benchmarking frameworks like PhishBench have been developed to evaluate and compare existing features for phishing detection, taking into account dataset characteristics and the need for new features and techniques to combat evolving attacks [59]. Anomaly detection approaches, particularly using deep learning, have shown promise in detecting phishing web pages, addressing the challenges of training samples and novel phishing attacks . Multiple techniques are being implemented to mitigate specific phishing attacks, including detection, offensive defense, correction, and prevention, highlighting the need for a holistic approach to phishing mitigation [60].

3. Proposed Method

The proposed model is a Sequential neural network architecture, specifically a Convolutional Neural Network (CNN), designed for binary classification tasks. In this study, convolutional layers are used to extract essential features from the input data. Convolution is performed by sliding a 3-sized kernel over the entire input data. This model consists of three convolutional layers with 128, 256, and 512 filters, respectively, each followed by a Rectified Linear Unit (ReLU) activation function, in equations (1)

$$Z^{[l]} = W^{[l]} * A^{[l-1]} + b^{[l]} \quad (1)$$

where $Z^{[l]}$ is the convolution output, $W^{[l]}$ is the weights, $A^{[l-1]}$is the output from the previous layer, and $b^{[l]}$is the bias.

We also adopt Max-pooling layers to reduce the dimensionality of the data by taking the maximum value within a sliding window after the convolution operation, in equations (2)

$$Y^{[l]}[i] = max\left(A^{[l]}[i: i + f]\right) \quad (2)$$

where $Y^{[l]}$ is the max-pooling output, $A^{[l]}$ is the input, $[i]$is the index, and $[f]$ is the window size. Dropout layers are incorporated to prevent overfitting by randomly deactivating some units during training, in equations (3)

$$[A^{\{[l]\}}_{\{dropout\}} = A^{\{[l]\}} * D^{\{[l]\}}] \quad (3)$$

where $A^{\{[l]\}}$ is the output before dropout, and $D^{\{[l]\}}$ is the dropout matrix containing values of 0 (inactive) or 1 (active).

This study utilizes GlobalAveragePooling1D layer to average the output of the previous convolutional layers to obtain a single value per filter, in equations (4)

$$\text{GAP}[i] = \frac{1}{n}\sum_{j=1}^{n} X[i,j] \quad (4)$$

Where GAP$[i]$ is the average value per filter, $X[i,j]$ is the output from the previous layer, and n is the number of elements in the filter. At the final layer, dense (fully connected) layers are utilized to combine the features extracted by the convolutional layers and perform binary classification. This model has one hidden layer with 128 units and a ReLU activation function, and one output unit with a sigmoid activation function, in equations (5).

$$[Z^{[l]} = W^{[l]} \cdot A^{[l-1]} + b^{[l]}] \quad (5)$$

where $Z^{[l]}$ is the dense layer output, $W^{[l]}$ is the weights, $A^{[l-1]}$ is the output from the previous layer, and $b^{[l]}$ is the bias. The model is compiled using binary cross-entropy loss and an optimizer. The training process consists of 200 epochs with a batch size of 64. Throughout training, monitoring is performed using the `val_checkpoint` callback. The goal of this model is to classify input data into two classes (0 or 1) based on learning from the training data.

4. Experimental Setup

A. Dataset

The dataset utilized in this research is the Phishing Detection Dataset obtained from Kaggle. This dataset comprises 14,093 records and features 29 diverse attributes. The primary objective of this study is to perform detection between two main categories: "Legitimate" (legitimate websites) and "Phishing" (suspicious websites). This dataset serves as the foundation for training and testing the phishing detection model in this research.

B. Preprocessing

In the context of this research, several preprocessing steps have been employed to prepare the dataset for effective machine learning model training. Initially, irrelevant features were removed from the dataset to streamline the data. Following that, rows containing missing values (NAN) were eliminated to ensure data consistency. Additionally, string-based data was converted into integer format to make it compatible with the machine learning model. Further, normalization techniques were applied to the 'age' feature to scale its values appropriately. Lastly, the dataset was divided into two subsets: 80% for training the model and 20% for testing its performance, ensuring a robust evaluation process for the phishing detection model.

C. Comparison of Methods

This study leverages the CNN algorithm as its primary approach for phishing detection. Additionally, it employs several baseline algorithms, including SVM, Decision Tree, KNN, Gaussian Naive Bayes, and Gradient Boost, to conduct a comparative analysis of their performance. By utilizing this ensemble of algorithms, the research aims to evaluate the effectiveness of CNN in detecting phishing attacks and compare it to the baseline algorithms, providing valuable insights into the potential advantages of CNN-based solutions in the realm of cybersecurity.

D. Training and Evaluation

In this research, the model is trained using 200 epochs and a batch size of 64. The performance of the model is evaluated using the Confusion Matrix and Classification Report. Confusion Matrix is a table used to measure the performance of a classification model. It consists of four main components: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN).

$$Confusion\ Matrix = \begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix} \tag{6}$$

Classification Report is a report that contains various classification evaluation metrics such as Precision, Recall, F1-Score, and Accuracy. These metrics are calculated using the components of the Confusion Matrix.

$$\begin{aligned} Precision &= \{TP\}\{(TP\ +\ FP\} \\ Recall &= \{TP\}\{(TP\ +\ FN\} \\ F1-Score &\frac{\{2 *\ Precision\ *\ Recall\}}{(Precision + Recall)} \\ Accuracy\ &= \{\vdash(TP + TN\dashv)\}\{(TP + TN + FP + FN\} \end{aligned} \tag{7}$$

This research can mathematically evaluate the performance of the trained model in classifying data.

5. Result and Analysis

A. Training Process

In this research, we present two graphs that illustrate the performance of the trained model using the CNN algorithm. The first graph (Figure 1) displays the training accuracy curve (blue line) and the validation accuracy curve (orange line) throughout the training process. This graph provides insights into how well the model comprehends and classifies the data correctly.

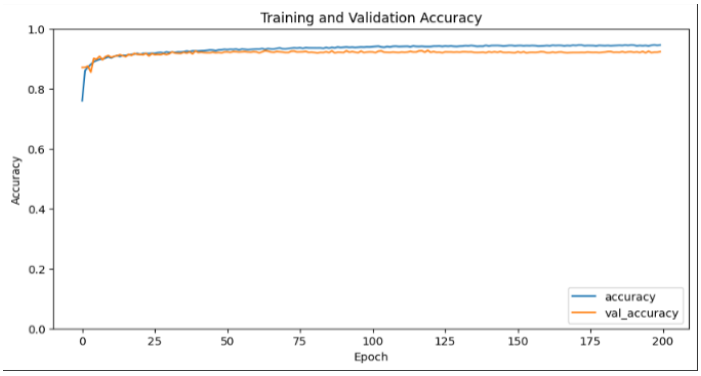

Figure 1. Training and Validation Accuracy

The second graph (Figure 2) visualizes the training loss curve (blue line) and the validation loss curve (orange line). Loss is a metric that measures how close the model approaches the correct results, and this graph helps in understanding to what extent the model optimizes its objective function during training. Analyzing these two graphs will provide a deeper understanding of the capabilities of the CNN model used in phishing detection.

Figure 2. Training and Validation Loss

B. Model Performance

Table 1, which presents the confusion matrix of several algorithms evaluated in this study, including CNN, SVM, Decision Tree, KNN, GNB, and GBoost. This table provides a comprehensive overview of the performance of each algorithm in classifying instances as either phishing or legitimate. The analysis from this confusion matrix will help us understand the extent of each algorithm's capability in detecting phishing attacks and identifying accurate outcomes.

Table 1.Confusion Matrix				
Algorithm	**TP**	**FP**	**FN**	**TN**
CNN	1313	97	107	1302
SVM	1254	156	90	1319
Decision Tree	1253	157	167	1242
KNN	1302	108	143	1266
GNB	1410	0	1372	37
GBoost	1249	161	102	1307

Table 2, which presents the classification report metrics of various algorithms assessed in this study, including CNN, SVM, Decision Tree, KNN, GNB, and GBoost. Table 2 provides a comprehensive evaluation of each algorithm's performance in terms of accuracy, precision, recall, and f1-score. These metrics offer valuable insights into the algorithms' abilities to classify phishing and legitimate instances effectively. By examining the classification report, we can gain a deeper understanding of the strengths and weaknesses of each algorithm in the context of phishing detection.

Table 2. Classification Report					
Algorithm	**Accuracy**	**Class**	**Precision**	**Recall**	**f1-score**
CNN	0.93	legitimate	0.92	0.93	0.93
		phishing	0.93	0.92	0.93
SVM	0.91	legitimate	0.93	0.89	0.91
		phishing	0.89	0.94	0.91
Decision Tree	0.89	legitimate	0.88	0.89	0.89
		phishing	0.89	0.88	0.88
KNN	0.91	legitimate	0.90	0.92	0.91
		phishing	0.92	0.90	0.91
GNB	0.51	legitimate	0.51	1.00	0.67
		phishing	1.00	0.03	0.05
GBoost	0.91	legitimate	0.92	0.89	0.90
		phishing	0.89	0.93	0.91

Table 2 presents the classification performance of six machine learning algorithms for phishing website detection. Among all evaluated models, the Convolutional Neural Network (CNN) achieved the highest overall accuracy of 0.93. The CNN also produced balanced performance for both classes, obtaining a precision, recall, and F1-score of approximately 0.93 for legitimate and phishing websites. These results indicate that the CNN effectively learned discriminative patterns from the input features and provided the most reliable classification performance. In comparison, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Gradient Boosting (GBoost) each achieved an accuracy of 0.91, demonstrating competitive performance with only a slight decrease compared with CNN. Meanwhile, the Decision Tree classifier obtained an accuracy of 0.89, indicating satisfactory but relatively lower predictive capability.

A detailed analysis of the class-level metrics shows that SVM, KNN, and GBoost maintained a balanced trade-off between precision and recall for both legitimate and

phishing classes. SVM achieved high precision (0.93) for legitimate websites and high recall (0.94) for phishing websites, indicating its effectiveness in minimizing false negatives. Similarly, KNN produced nearly identical precision and recall values for both classes, resulting in a consistent F1-score of 0.91. GBoost also demonstrated stable classification performance with F1-scores of 0.90 for legitimate websites and 0.91 for phishing websites. In contrast, the Decision Tree classifier exhibited slightly lower precision and recall values, suggesting that single-tree models may have limited capability in capturing the complex decision boundaries required for phishing website classification.

The Gaussian Naïve Bayes (GNB) classifier produced the weakest performance among all evaluated methods, achieving an overall accuracy of only 0.51. Although GNB correctly identified almost all legitimate websites with a recall of 1.00, it detected only 3% of phishing websites, resulting in an F1-score of 0.05 for the phishing class. This substantial performance gap indicates that the independence assumption of GNB does not adequately model the complex relationships among phishing website features. Overall, the experimental results demonstrate that deep learning and ensemble-based approaches, particularly CNN and GBoost, provide more accurate and balanced phishing detection than conventional probabilistic classifiers, making them more suitable for practical cybersecurity applications.

6. Conclusion

This paper developed and evaluated multiple machine learning models for phishing website detection, including Convolutional Neural Network (CNN), Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbor (KNN), Gaussian Naïve Bayes (GNB), and Gradient Boosting (GBoost). We first analyzed the learning behavior of the CNN model through training and validation accuracy and loss curves. The results showed that the training process converged successfully, with the validation curves closely following the training curves. This behavior indicates stable learning, good generalization capability, and no significant evidence of overfitting. These findings demonstrate that the CNN effectively learned discriminative representations from the phishing website dataset and established a robust foundation for accurate website classification.

The experimental evaluation further confirmed that the CNN achieved the best overall performance among all evaluated algorithms. The proposed CNN model obtained an accuracy of 93%, with balanced precision, recall, and F1-score of approximately 0.93 for both legitimate and phishing classes. SVM, KNN, and GBoost also produced competitive results, each achieving 91% accuracy while maintaining a good balance between precision and recall. In contrast, the Decision Tree model achieved a slightly lower accuracy of 89%, indicating that a single-tree structure was less effective in capturing the complex characteristics of phishing websites. Meanwhile, GNB produced the weakest performance with only 51% accuracy because its feature independence assumption failed to model the complex relationships among phishing website attributes. These results demonstrate that deep learning and ensemble-based approaches provide more reliable phishing detection than conventional probabilistic classifiers.

Overall, this paper demonstrates that CNN is the most effective approach for phishing website detection in the evaluated dataset because it provides the highest accuracy and the most balanced classification performance across both website categories. We utilize comprehensive performance metrics, including confusion matrices, accuracy, precision, recall, and F1-score, to validate the effectiveness of each algorithm objectively. The findings indicate that deep learning can significantly improve phishing detection performance and strengthen practical cybersecurity applications by reducing both false positives and false negatives. Future work may investigate more advanced deep learning

architectures, such as Vision Transformers (ViT), hybrid CNN-transformer models, or Graph Neural Networks, while incorporating feature selection, explainable artificial intelligence (XAI), and larger real-world phishing datasets to further improve detection accuracy, model interpretability, and robustness against evolving phishing attacks.

Acknowledgments

This paper is conducted in the Department of Informatics, Universitas Respati Yogyakarta, Indonesia.

Data Availability

The datasets come from the Kaggle website https://www.kaggle.com/datasets/sandunabeysooriya/phishing-detection-dataset?datasetId=1291399)

References

[1] M. Hussain, C. Cheng, R. Xu, and M. Afzal, “CNN-Fusion: An effective and lightweight phishing detection method based on multi-variant ConvNet,” Inf. Sci., vol. 631, pp. 328–345, Jun. 2023, doi: 10.1016/j.ins.2023.02.039.

[2] Z. Alshingiti, R. Alaqel, J. Al-Muhtadi, Q. E. Haq, K. Saleem, and M. H. Faheem, “A Deep Learning-Based Phishing Detection System Using CNN, LSTM, and LSTM-CNN,” Electronics, vol. 12, no. 1, 2023, doi: 10.3390/electronics12010232.

[3] X. Xiao et al., “Phishing websites detection via CNN and multi-head self-attention on imbalanced datasets,” Comput. Secur., vol. 108, p. 102372, Sep. 2021, doi: 10.1016/j.cose.2021.102372.

[4] S. Srivastava and S. K. Gupta, “Phishing Detection Techniques: A Comparative Study,” in 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Sep. 2021, pp. 1–6. doi: 10.1109/ICRITO51393.2021.9596093.

[5] J. A. Bukhsh, “Guidelines for Developers and Recommendations for Users to Mitigate Phishing Attacks: An Interdisciplinary Research Approach,” in Research Challenges in Information Science: Information Science and the Connected World, S. Nurcan, A. L. Opdahl, H. Mouratidis, and A. Tsohou, Eds., Cham: Springer Nature Switzerland, 2023, pp. 632–640.

[6] A. K. Jain and B. B. Gupta, “A survey of phishing attack techniques, defence mechanisms and open research challenges,” Enterp. Inf. Syst., vol. 16, no. 4, pp. 527–565, Apr. 2022, doi: 10.1080/17517575.2021.1896786.

[7] P. K. Sahoo, “Data mining a way to solve Phishing Attacks,” in 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), Mar. 2018, pp. 1–5. doi: 10.1109/ICCTCT.2018.8550910.

[8] R. M. Mohammad, F. Thabtah, and L. McCluskey, “Predicting phishing websites based on self-structuring neural network,” Neural Comput. Appl., vol. 25, no. 2, pp. 443–458, Aug. 2014, doi: 10.1007/s00521-013-1490-z.

[9] B. Chakraborty et al., “Detection of Rice Blast Disease (Magnaporthe grisea) Using Different Machine Learning Techniques,” Int. J. Environ. Clim. Change, vol. 13, no. 8, pp. 2256–2264, Jun. 2023, doi: 10.9734/ijecc/2023/v13i82190.

[10] D. Dalalah and O. M. A. Dalalah, “The false positives and false negatives of generative AI detection tools in education and academic research: The case of ChatGPT,” Int. J. Manag. Educ., vol. 21, no. 2, p. 100822, Jul. 2023, doi: 10.1016/j.ijme.2023.100822.

[11] R. Bold, H. Al-Khateeb, and N. Ersotelos, “Reducing False Negatives in Ransomware Detection: A Critical Evaluation of Machine Learning Algorithms,” Appl. Sci., vol. 12, no. 24, 2022, doi: 10.3390/app122412941.

[12] T. Choudhary, S. Mhapankar, R. Bhddha, A. Kharuk, and Dr. Rohini Patil, “A Machine Learning Approach for Phishing Attack Detection,” J. Artif. Intell. Technol., vol. 3, no. 3, pp. 108–113, May 2023, doi: 10.37965/jait.2023.0197.
[13] T. Choudhary, S. Mhapankar, R. Bhddha, A. Kharuk, and Dr. Rohini Patil, “A Machine Learning Approach for Phishing Attack Detection,” J. Artif. Intell. Technol., vol. 3, no. 3, pp. 108–113, May 2023, doi: 10.37965/jait.2023.0197.
[14] S. R. Abdul Samad et al., “Analysis of the Performance Impact of Fine-Tuned Machine Learning Model for Phishing URL Detection,” Electronics, vol. 12, no. 7, 2023, doi: 10.3390/electronics12071642.
[15] A. H. Aljammal, S. taamneh, A. Qawasmeh, and H. Bani Salameh, “Machine Learning Based Phishing Attacks Detection Using Multiple Datasets,” Int. J. Interact. Mob. Technol. IJIM, vol. 17, no. 05, pp. 71–83, Mar. 2023, doi: 10.3991/ijim.v17i05.37575.
[16] D. T. Mosa, M. Y. Shams, A. A. Abohany, E.-S. M. El-kenawy, and M. Thabet, “Machine Learning Techniques for Detecting Phishing URL Attacks,” Comput. Mater. Contin., vol. 75, no. 1, 2023, doi: 10.32604/cmc.2023.036422.
[17] M. Aljabri and S. Mirza, “Phishing Attacks Detection using Machine Learning and Deep Learning Models,” in 2022 7th International Conference on Data Science and Machine Learning Applications (CDMA), Mar. 2022, pp. 175–180. doi: 10.1109/CDMA54072.2022.00034.
[18] A. Hannousse and S. Yahiouche, “Towards benchmark datasets for machine learning based website phishing detection: An experimental study,” Eng. Appl. Artif. Intell., vol. 104, p. 104347, Sep. 2021, doi: 10.1016/j.engappai.2021.104347.
[19] L. Tang and Q. H. Mahmoud, “A Survey of Machine Learning-Based Solutions for Phishing Website Detection,” Mach. Learn. Knowl. Extr., vol. 3, no. 3, pp. 672–694, 2021, doi: 10.3390/make3030034.
[20] S. Zaman, S. M. Uddin Deep, Z. Kawsar, M. Ashaduzzaman, and A. I. Pritom, “Phishing Website Detection Using Effective Classifiers and Feature Selection Techniques,” in 2019 2nd International Conference on Innovation in Engineering and Technology (ICIET), Dec. 2019, pp. 1–6. doi: 10.1109/ICIET48527.2019.9290554.
[21] T. Choudhary, S. Mhapankar, R. Bhddha, A. Kharuk, and Dr. Rohini Patil, “A Machine Learning Approach for Phishing Attack Detection,” J. Artif. Intell. Technol., vol. 3, no. 3, pp. 108–113, May 2023, doi: 10.37965/jait.2023.0197.
[22] A. H. Aljammal, S. taamneh, A. Qawasmeh, and H. Bani Salameh, “Machine Learning Based Phishing Attacks Detection Using Multiple Datasets,” Int. J. Interact. Mob. Technol. IJIM, vol. 17, no. 05, pp. 71–83, Mar. 2023, doi: 10.3991/ijim.v17i05.37575.
[23] S. Alnemari and M. Alshammari, “Detecting Phishing Domains Using Machine Learning,” Appl. Sci., vol. 13, no. 8, 2023, doi: 10.3390/app13084649.
[24] Z. Alshingiti, R. Alaqel, J. Al-Muhtadi, Q. E. Haq, K. Saleem, and M. H. Faheem, “A Deep Learning-Based Phishing Detection System Using CNN, LSTM, and LSTM-CNN,” Electronics, vol. 12, no. 1, 2023, doi: 10.3390/electronics12010232.
[25] D. T. Mosa, M. Y. Shams, A. A. Abohany, E.-S. M. El-kenawy, and M. Thabet, “Machine Learning Techniques for Detecting Phishing URL Attacks,” Comput. Mater. Contin., vol. 75, no. 1, 2023, doi: 10.32604/cmc.2023.036422.
[26] A. Safi and S. Singh, “A systematic literature review on phishing website detection techniques,” J. King Saud Univ. - Comput. Inf. Sci., vol. 35, no. 2, pp. 590–611, Feb. 2023, doi: 10.1016/j.jksuci.2023.01.004.
[27] R. R. Patil et al., “Machine learning approach for phishing website detection : A literature survey,” J. Discrete Math. Sci. Cryptogr., vol. 25, no. 3, pp. 817–827, Apr. 2022, doi: 10.1080/09720529.2021.2016224.
[28] M. Sánchez-Paniagua, E. Fidalgo, E. Alegre, and R. Alaiz-Rodríguez, “Phishing websites detection using a novel multipurpose dataset and web technologies features,” Expert Syst. Appl., vol. 207, p. 118010, Nov. 2022, doi: 10.1016/j.eswa.2022.118010.
[29] Yi He, “Phishing websites detection by machine learning and deep learning techniques,” presented at the Proc.SPIE, Nov. 2022, p. 123012C. doi: 10.1117/12.2644665.
[30] S. Jain and C. Gupta, “A Support Vector Machine Learning Technique for Detection of Phishing Websites,” in 2023 6th International Conference on Information Systems and

Computer Networks (ISCON), Mar. 2023, pp. 1–6. doi: 10.1109/ISCON57294.2023.10111968.
[31] A. Hannousse and S. Yahiouche, “Towards benchmark datasets for machine learning based website phishing detection: An experimental study,” Eng. Appl. Artif. Intell., vol. 104, p. 104347, Sep. 2021, doi: 10.1016/j.engappai.2021.104347.
[32] M. H. Alkawaz, S. J. Steven, A. I. Hajamydeen, and R. Ramli, “A Comprehensive Survey on Identification and Analysis of Phishing Website based on Machine Learning Methods,” in 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), Apr. 2021, pp. 82–87. doi: 10.1109/ISCAIE51753.2021.9431794.
[33] M. D. Bhagwat, P. H. Patil, and T. S. Vishawanath, “A Methodical Overview on Detection, Identification and Proactive Prevention of Phishing Websites,” in 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Feb. 2021, pp. 1505–1508. doi: 10.1109/ICICV50876.2021.9388441.
[34] E. Gandotra and D. Gupta, “An Efficient Approach for Phishing Detection using Machine Learning,” in Multimedia Security: Algorithm Development, Analysis and Applications, K. J. Giri, S. A. Parah, R. Bashir, and K. Muhammad, Eds., Singapore: Springer Singapore, 2021, pp. 239–253. doi: 10.1007/978-981-15-8711-5_12.
[35] G. Matošević, J. Dobša, and D. Mladenić, “Using Machine Learning for Web Page Classification in Search Engine Optimization,” Future Internet, vol. 13, no. 1, 2021, doi: 10.3390/fi13010009.
[36] R. Zaimi, M. Hafidi, and M. Lamia, “A deep learning approach to detect phishing websites using CNN for privacy protection,” Intell. Decis. Technol., vol. 17, no. 3, pp. 713–728, 2023, doi: 10.3233/IDT-220307.
[37] S. Paliath, M. A. Qbeitah, and M. Aldwairi, “PhishOut: Effective Phishing Detection Using Selected Features,” in 2020 27th International Conference on Telecommunications (ICT), Oct. 2020, pp. 1–5. doi: 10.1109/ICT49546.2020.9239589.
[38] A. Das, S. Baki, A. El Aassal, R. Verma, and A. Dunbar, “SoK: A Comprehensive Reexamination of Phishing Research From the Security Perspective,” IEEE Commun. Surv. Tutor., vol. 22, no. 1, pp. 671–708, Firstquarter 2020, doi: 10.1109/COMST.2019.2957750.
[39] J. Kamiri and G. Mariga, “Research Methods in Machine Learning: A Content Analysis,” Int. J. Comput. Inf. Technol.-0764, vol. 10, no. 2, Mar. 2021, doi: 10.24203/ijcit.v10i2.79.
[40] A. Abuzuraiq, M. Alkasassbeh, and M. Almseidin, “Intelligent Methods for Accurately Detecting Phishing Websites,” in 2020 11th International Conference on Information and Communication Systems (ICICS), Apr. 2020, pp. 085–090. doi: 10.1109/ICICS49469.2020.239509.
[41] I. F. Ferdiansyah and Wella, “Comparative Study of Hoax Detection using Support Vector Machine Algorithm, Naive Bayes, Random Forest and K-Nearest Neighbor,” in 2022 13th International Conference on Information and Communication Technology Convergence (ICTC), Oct. 2022, pp. 1123–1126. doi: 10.1109/ICTC55196.2022.9952630.
[42] A. H. Aljammal, S. taamneh, A. Qawasmeh, and H. Bani Salameh, “Machine Learning Based Phishing Attacks Detection Using Multiple Datasets,” Int. J. Interact. Mob. Technol. IJIM, vol. 17, no. 05, pp. 71–83, Mar. 2023, doi: 10.3991/ijim.v17i05.37575.
[43] Y. Almaghthawi, I. Ahmad, and F. E. Alsaadi, “Performance Analysis of Feature Subset Selection Techniques for Intrusion Detection,” Mathematics, vol. 10, no. 24, 2022, doi: 10.3390/math10244745.
[44] A. Hannousse and S. Yahiouche, “Towards benchmark datasets for machine learning based website phishing detection: An experimental study,” Eng. Appl. Artif. Intell., vol. 104, p. 104347, Sep. 2021, doi: 10.1016/j.engappai.2021.104347.
[45] X. Xiao et al., “Phishing websites detection via CNN and multi-head self-attention on imbalanced datasets,” Comput. Secur., vol. 108, p. 102372, Sep. 2021, doi: 10.1016/j.cose.2021.102372.
[46] S. Paliath, M. A. Qbeitah, and M. Aldwairi, “PhishOut: Effective Phishing Detection Using Selected Features,” in 2020 27th International Conference on Telecommunications (ICT), Oct. 2020, pp. 1–5. doi: 10.1109/ICT49546.2020.9239589.
[47] T. Choudhary, S. Mhapankar, R. Bhddha, A. Kharuk, and Dr. Rohini Patil, “A Machine Learning Approach for Phishing Attack Detection,” J. Artif. Intell. Technol., vol. 3, no. 3, pp. 108–113, May 2023, doi: 10.37965/jait.2023.0197.
[48] T. Choudhary, S. Mhapankar, R. Bhddha, A. Kharuk, and Dr. Rohini Patil, “A Machine Learning Approach for Phishing Attack Detection,” J. Artif. Intell. Technol., vol. 3, no. 3, pp. 108–113, May 2023, doi: 10.37965/jait.2023.0197.

[49] A. Abuzuraiq, M. Alkasassbeh, and M. Almseidin, “Intelligent Methods for Accurately Detecting Phishing Websites,” in 2020 11th International Conference on Information and Communication Systems (ICICS), Apr. 2020, pp. 085–090. doi: 10.1109/ICICS49469.2020.239509.
[50] R. Zaimi, M. Hafidi, and M. Lamia, “A deep learning approach to detect phishing websites using CNN for privacy protection,” Intell. Decis. Technol., vol. 17, no. 3, pp. 713–728, 2023, doi: 10.3233/IDT-220307.
[51] Z. Alshingiti, R. Alaqel, J. Al-Muhtadi, Q. E. Haq, K. Saleem, and M. H. Faheem, “A Deep Learning-Based Phishing Detection System Using CNN, LSTM, and LSTM-CNN,” Electronics, vol. 12, no. 1, 2023, doi: 10.3390/electronics12010232.
[52] N. J. Sinthiya, T. A. Chowdhury, and A. B. Haque, “Incorporating Machine Learning Algorithms to Detect Phishing Websites,” in 2022 International Conference on ICT for Smart Society (ICISS), Aug. 2022, pp. 1–5. doi: 10.1109/ICISS55894.2022.9915211.
[53] S. Das Guptta, K. T. Shahriar, H. Alqahtani, D. Alsalman, and I. H. Sarker, “Modeling Hybrid Feature-Based Phishing Websites Detection Using Machine Learning Techniques,” Ann. Data Sci., Mar. 2022, doi: 10.1007/s40745-022-00379-8.
[54] S. Salloum, T. Gaber, S. Vadera, and K. Shaalan, “Phishing Website Detection from URLs Using Classical Machine Learning ANN Model,” in Security and Privacy in Communication Networks, J. Garcia-Alfaro, S. Li, R. Poovendran, H. Debar, and M. Yung, Eds., Cham: Springer International Publishing, 2021, pp. 509–523.
[55] B. B. Gupta, K. Yadav, I. Razzak, K. Psannis, A. Castiglione, and X. Chang, “A novel approach for phishing URLs detection using lexical based machine learning in a real-time environment,” Comput. Commun., vol. 175, pp. 47–57, Jul. 2021, doi: 10.1016/j.comcom.2021.04.023.
[56] B. B. Gupta, K. Yadav, I. Razzak, K. Psannis, A. Castiglione, and X. Chang, “A novel approach for phishing URLs detection using lexical based machine learning in a real-time environment,” Comput. Commun., vol. 175, pp. 47–57, Jul. 2021, doi: 10.1016/j.comcom.2021.04.023.
[57] E. Gandotra and D. Gupta, “An Efficient Approach for Phishing Detection using Machine Learning,” in Multimedia Security: Algorithm Development, Analysis and Applications, K. J. Giri, S. A. Parah, R. Bashir, and K. Muhammad, Eds., Singapore: Springer Singapore, 2021, pp. 239–253. doi: 10.1007/978-981-15-8711-5_12.
[58] D. N. Quang, A. Selamat, and O. Krejcar, “Recent Research on Phishing Detection Through Machine Learning Algorithm,” in Advances and Trends in Artificial Intelligence. Artificial Intelligence Practices, H. Fujita, A. Selamat, J. C.-W. Lin, and M. Ali, Eds., Cham: Springer International Publishing, 2021, pp. 495–508.
[59] A. El Aassal, S. Baki, A. Das, and R. M. Verma, “An In-Depth Benchmarking and Evaluation of Phishing Detection Research for Security Needs,” IEEE Access, vol. 8, pp. 22170–22192, 2020, doi: 10.1109/ACCESS.2020.2969780.
[60] L. Ouyang and Y. Zhang, “Phishing Web Page Detection with Semi-Supervised Deep Anomaly Detection,” in Security and Privacy in Communication Networks, J. Garcia-Alfaro, S. Li, R. Poovendran, H. Debar, and M. Yung, Eds., Cham: Springer International Publishing, 2021, pp. 384–393.