International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME003
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Improving Phishing Detection Accuracy Through Feature Selection and Deep Learning Model Integration

MEKALA MANOJ KUMAR
Department of computer science and engineering,
Vemu Institute of technology
kothakota, chittor district,
Andhra Pradesh 517112,
Mailid: mekalamanoj7359@gmail.com

Mr. K. Niranjan
Assistant Professor
Vemu Institute of Technology,
Dept of CSE
Mail id- kalikiriniranjan@gmail.com

Abstraction: Phishing attacks are a persistent cybersecurity threat, and thieves are able to steal important information by exploiting people's trust. In this paper, we present an enhanced phishing detection framework, which is based on the combination of advanced feature selection methods and Machine Learning and Deep Learning algorithm. The paper leverages the use of labeled dataset with each instance being either legitimate or phishing and introduces the performance of using different models Graph Convolutional Networks (GCN), TabTransformer, Autoencoders, Feedforward Neural Networks (FNN) and Deep Neural Networks (DNN). Finally, the feature selection process should be optimized to help improve the model accuracy, decrease the computational overhead, and improve the generalization. The solution has been coded in python and released in the form of a flask web application with a simple HTML and CSS based user interface. Our experiment results support the conclusion that a deep learning architecture in combination with a proper feature selection can facilitate a better performance and robust phishing detection mechanism. This is a real-life implementation model which can be used to achieve a scalable and effective phishing control.

Keywords: DNN, FNN, Autoencoder, GCN, TabTransformer, Feature generation, Phishing Detection, Flask, Cybersecurity.

I. INTRODUCTION

Phishing attacks remain a high cybersecurity risk and are a threat to privacy and security of people and organisations in the global front. In essence, brute force attacks are automated guesses that are utilized in impersonation of legitimate organizations to whom the customers receive emails with an aim to access their personal information including user IDs and passwords, account financial information or other personal identifiers. Phishing tools have evolved further than just email phishing attacks, they have started to more intelligently employ more advanced forms of social engineering and this makes it harder to know what is genuine communication or an attack. The complexity of phishing attacks is getting elevated day by day, there is a demand for a complex detection system to detect phishing attacks in real-time, and block the relevant traffic. The cert-based approach to signing in is no longer capable of quick enough response to the dynamic nature of phishing. However, cybersecurity industry has been undertaking ML and DL processes to develop more effective solutions to combat attacks by sifting through data piles and learning on stimuli including phishing attacks. Phishing attacks exist in various categories, which are email phishing, social media phishing, voice phishing, text message phishing and spear-phishing. These risks will typically be email messages, masquerading as installers from reputable organizations such as financial institutions, governments and e-commerce channels. The authors of these attacks are using sentences analysis methods like inserts websites that resemble these johns and jans of the cyber world, and these links to websites which use the victims to transmit private information. The danger of phishing is now well known to the public, and is, nevertheless, one of the most effective methods used by cybercriminals, not the least because its success is predicated on human psychology and is therefore, not only predicated on technological vulnerabilities.

Cybersecurity attackers have resorted to machine learning as a subdivision of artificial intelligence (AI),

whereby a system is programmed to learn information and make its own decisions without further instructions to combat such advanced attacks. Machine learning can be trained to identify trends and abnormalities in data and therefore, can be highly useful at detecting phishing attacks. With the emphasis on different properties of the cyber clues such as email content, URL, sender information, website properties, and so on, machine learning algorithms are capable of identifying whether a particular entity is a legitimate one or a Phishing attack. Specifically, one of the subfields of machine learning, deep learning, is particularly effective in finding complex trends on large amounts of data that is exceptionally useful in phishing detection because the patterns that should be discovered are delicate and cannot be detected through traditional means.

Feature selection is an important factor in the performance of phishing detection model. As an effect of buying bulk of data, it's important to have the most significant and relevant features used for enhanced accuracy and effectual model and calculation intricacy. Feature selection algorithm has been applied to identify training attributes which are most significant in the raw data set regarding the model significance to differentiate the real user and phishing activity. Some of these features will consist of features of URLs, content of e-mails, metadata and domain information, and even user behavior itself. Feature selection can help to make the model more efficient but it can also help to avoid overfitting, in other words when the model is too complex and will capture noise in the data as opposed to the true underlying pattern. For instance, useful data points such as URL length, presence of special characters, HTTPS lookup of website domain reputation, HTS encryption, etc can be used to identify phishing sites. Similarly, the email characteristics like suspicious subject line, hyperlinks and message tone, etc., can be used to identify phishing emails.

More specifically, the machine learning and deep learning models that are complementary by the feature selection methods can also be used to improve the phishing detection even more. These systems can be made to be more general, more accurate and efficient by optimising the characteristics of the features used in training the system to design an accurate, robust and efficient system capable of solving the variety of phishing systems. In this regard, machine learning models have proven to be very promising. Decision trees and random forests are most suited to the task of processing structured data whose patterns can be expressed in the form of a sequence of if-then-else rules. However, the SVM method is quite effective in separating data classes based on a hyperplane, which maximizes the separation between the classes by maximizing the distance between the classes. Particularly, multi-layer neural networks have been the most powerful tools in detecting phishing by automatically learning highly non-linear correlations in the big data such that the model learns to detect phishing even without having a definition of that phishing.

II. RELATED WORK

Researchers have studied machine learning (ML) models for phishing detection since the last few years while focusing on supervised learning algorithms that include DT and SVM. The methods classify URLs and emails into two categories which are phishing and legitimate. The research shows that automated detection systems need implementation because they will improve threat detection capabilities through real-time monitoring while reducing online fraud risks [1].

The research paper provides multiple phishing detection techniques which combine machine learning algorithms with Decision Trees and Random Forest and heuristic methods. The research studies various methods to combat phishing attacks while demonstrating their success in defending against spear phishing attacks smishing attacks and email phishing attacks[2].

The complete study identifies different types of phishing attacks which include spear phishing whaling and clone phishing as its main focus. The study examines three types of countermeasures which include URL detection methods and content analysis techniques and machine learning detection systems. The paper demonstrates the requirement for adaptive detection systems which protect against evolving phishing attack techniques [3], [4].

The research investigates various machine learning techniques which identify phishing websites by analyzing both URLs and website content. The study evaluates SVM Decision Trees and Logistic Regression through standard datasets to show that feature selection methods are essential for achieving better model accuracy [5].

Researchers test machine learning algorithms through Decision Trees and Random Forest methods to identify phishing websites. The research investigates classification methods which URL links and HTML content link together while showing that continuous

model development is necessary to combat emerging phishing threats [6], [7].

The research demonstrates that feature engineering works as a critical element for identifying phishing attacks. The research develops machine learning models from three types of data which include email features and URL characteristics and domain reputation information. The researchers study feature selection methods which will assist them in identifying essential features for their training model development [8].

The researchers explored deep learning models in their research that comprises of Convolutional Neural Networks and Recurrent Neural Networks in the context of determining its effectiveness in identifying phishing attack. The models show that they can identify phishing attacks since it can automatically learn complex patterns of data without researchers having to create particular features that it uses [9], [10].

The study indicates that the approximate accuracy of the phishing detection systems can be enhanced through the use of ensemble machine learning models. The system achieves better performance through enhanced precision and recall rates by using multiple classifiers. The implementation provides the system with stronger protection against various phishing attacks [11], [12].

Researchers study deep learning techniques to identify phishing emails which they use in their investigation. The research paper achieves its phishing email detection demonstration through the combination of feature selection and deep neural network models which deliver high detection accuracy [13].

The research study examines the method of using autoencoders to work with machine learning models to detect phishing attacks. The system involves the application of autoencoders to minimize dimensions that help the machine learning model focus on critical classification attributes.The study involves the use of machine learning techniques to detect phishing attacks that compromise the social media sites. The authors train their models with such features as the information of the sender and the content of the message and URL to identify the attempts of phishing [14], [15].

The research paper investigates how multi-layer neural networks function as a solution for detecting phishing attacks. The deep learning model is able to detect phishing websites and emails based on its capability to learn complicated patterns using unlabeled large datasets.The researchers experimented various deep learning models that comprise CNNs and LSTMs. The study indicates that the performance of deep learning models, in terms of accuracy and generalizations, is superior to the traditional ones [16], [17].

III. PROPOSED SYSTEM WORKFLOW

A. Data Collection

In the case of the phishing detection system, the model has been trained on a labeled dataset of the legitimate webpages and the phishing webpages and the results were tested to assess the work of model. There are several features borrowed from URLs, Emails, and web pages. Other variables like URL length, domain age, domain type and presence of HTTPS were added to try to extract possible phishing identifiers. Also, textual attributes like suspiciousness of word existence, hyperlinking and email headers were included for phishing email analysis. The phishing data set is a combination of phishing campaigns spanning from traditional email-based phishing attempts to advanced phishing on a higher level such as spear phishing attacks and smishing attacks. The collected data stream was sufficiently filtered to maintain balanced representation of both benign and malicious data such as to reflect realistic scenarios and provide sufficient training and evaluation of our model. The usage of real-world heterogeneous dataset provides the generalization ability of the detection system against various phishing strategies and websites. Training data and testing data set was a labeled data set containing phishing and legitimate web pages. The dataset contains a number of features related to both the URL and the webpage content, which should be very interesting for phishing detection.

B. Data Preprocessing

Data preprocessing is an important part of the preparation of the dataset to be used for the machine learning models. The raw data set contained a combination of structured features (e.g. URL and Domain information), and unstructured features (e.g. body of the email). The data first of all had some missing values and a large number of outliers were observed specially in the case of the numbers. Missing data had been treated by using statistical method of imputation and data used outliers were removed by using z-score or IQR methods to ensure consistent data. Categorical variables like type of domain (like '.com', '.org' etc) were represented using Label Encoding or One-Hot Encoding depending upon the feature type of

variable. For the numerical features MinMax scaling was employed to make the Uniformity and better speed of the Convergence of the model training. Feature selection techniques were also applied so the multicollinearity features could be removed, using correlation analysis, and only keep the most important features for model training. This preprocessing was done so that the data was well-prepared to be fed to the machine learning models, and there was no potential of introducing bias in these models because of some issues in the datat.

C. Model Development

The major part of the phishing detection system is based on the deep learning and machine learning models, and these models are built to categorize the website or emails as a genuine website or a phishing website. The model development process began from the comparative analysis of various ML algorithms like GCN, TabTransformer, Autoencoders, FNN and DNN. GCNs were specifically used because of their capacity of the relation between URLs, domains and emails in a graphical structure, which allows a better detection capacity for phishing attacks including interconnected entities. A transformer-based model that is termed as TabTransformer has been used to process the tabular data and learn any long range dependency between features for better classification performance. The autoencoders were used for anomaly detection, which is looking for patterns within the dataset that are not normal for website behavior. The choice of FNN and DNN models was made because of its ability to process and learn from high-dimensional datasets structured and unstructured data. A combination of these models were used in ensemble framework to better detect the accuracy and robustness. This hybrid approach can help the system to deal with the complexities on phishing detection, this includes different attack strategies and evolving phishing techniquest.

D. Model Evaluation

Model evaluation was performed by using a series of common classification metrics to evaluate the performance of proposed phishing detection framework. Metrics: Accuracy, Precision, Recall and F1-Score were calculated to evaluate the model performance. Precision provided the answer to the question of of the bodies found to be phishing websites, how many of them were truly phishing sites and accuracy can be considered to be how well the model classified phishing and legitimate websites. On the other hand, Recall was considered the metric to assess the detection of all the phishing activities performed by the model to guarantee a reduction of no phishing attacks missed. These two criteria were aggregated as a single evaluation metric through the F1, also referred to as F1-score, as the harmonic average of precision and recall. In the experiment results, an ensemble model consisting of different machine learning and deep learning techniques showed better results compared to the performance of the single model, which was a proper solution for phishing detection.

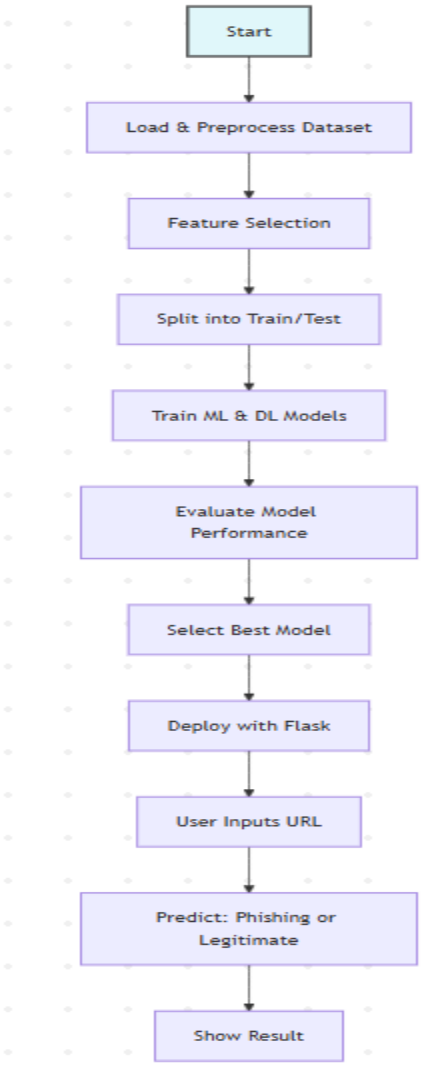


Figure 1 Project Work Flow

IV. METHODOLOGY

Graph Convolutional Networks (GCN):

GCN have been used as an effective tool to learn on graph-structured data where the relationships among the entities are critical to understand patterns. In particular, the use of GCNs is particularly well-suited to detecting phishing attacks, since phishing attacks are typically executed by way of network-based components like

malicious websites, senders and URLs that together form a developing network. The GCNs learn the dependencies between the nodes (e.g., URL, domain or IP addresses) in a graph, where each node represents an entity, and edges represent the relationships or the interaction that exist between them. This modeling of these relationships gives GCNs the ability to capture phishing data patterns that may not be immediately apparent for investigation using a traditional machine learning model that focuses on isolated features.

Graph convolutional networks (GCN) perform a convolution operation on a graph, meaning that information propagates along a graph. Each node in the graph sums up the features of its neighbours, which allows the model to learn high-level representations of nodes while retaining the information of the local structure of nodes. This message passing is repeated for a number of layers, so that the network sees long range dependencies and complex relationships. In the phishing detection case, for example, these dependencies can be used to detect phishing websites based on how they are linked to other suspect entities in the network, including phishing emails or malicious domains.

Autoencoders:

Autoencoders This category of neural network (Unsupervised neural networks) has primary applications in dimensionality reduction; learning features and detecting anomalies. They are made up of two big components which are encoder and decoder. The encoder are used to encode the input data to an encoding of the input data in lower dimension latent space and the decoder encode the original input by using the latent space representation. Auto encoder is designed to decrease the distance between the input and the reconstructed output that in turn learns important representations of the data.

Phishing is one of the situations that can use autoencoders and in this scenario; the autoencoders are able to acquire normal behaviour with genuine websites/ emails and any considerable deviation with normal behaviour would be acquired as suspicious phishing activities. The model is trained to recreate these normal inputs very accurately by being trained on the large data set of benign websites or a legal email traffic. On presenting the trained autoencoder with a new input in the form of a new phishing site or suspicious email, the trained autoencoder will produce reconstruction error (difference between original input and output of model) that will be more than normal data pointing to a potential anomaly.

Feedforward Neural Networks:

The most readily used and also well-known type of artificial neural networks is the Feedforward Neural Networks (FNN). In FNN, information flow is unidirectional and in a sequence i.e. input layer to hidden layers to the output layer and none of the feedback. It can be applied particularly to a case of supervised learning, where the data is labeled and the model is trained to map the input features into the predictions of the output. This network is multilayered: the data is being fed on the input layer, data are processed through hidden layer, and the final ones are formed on the output layer.

The FNNs are typically applied in the classification task, including phishing identifying. In this aspect, one can classify websites or emails as authentic or phishing with the help of FNNs using various types of input features e.g. URL features, email features or domain features. The hidden layers of the network learn these features either by training them and subsequently use the raw data to come up with meaningful representations that can be used to give accurate predictions. The neurons of the hidden layers apply mathematical activation function to the inputs to allow the network to be capable of imitating non-linear relationship in the data and this is particularly useful when working with complex patterns like those utilized by phishing attacks.

Deep Neural Networks (DNN)

DNNs is a type of neural networks which comprises of several layers connected to neurons, where each layer manipulates the input data by applying nonlinear transformations to it. DNNs have numerous hidden layers between the input and output unlike the traditional neural networks of a shallow structure allowing the DNNs to incorporate complex and hierarchical relationships in data. This higher-level representation enables DNNs to learn inherently complexity features of raw data, and so they are applicable to problems with the feature engineering process being challenging or time-intensive, like image recognition, speech recognition, and in phishing detection.

Concerning phishing detection, DNNs can be used to identify websites or emails as a legitimate source or malicious one based on different types of input features. Such characteristics may be URL characteristics,

domain details, webpage contents, or in the case of email, the body text of email. By subjecting the data to a sequence of hidden layers, the model learns the complex patterns and associations of the data that each layer pulls out an increasingly abstract representation of the data. Such a rich architecture enables DNNs to derive hidden trends in the data which could be evidence of phishing which could include minor variations in URL syntax and the unexpected occurrence of incongruent terminologies in the email message.

V. RESULTS

Autoencoders:

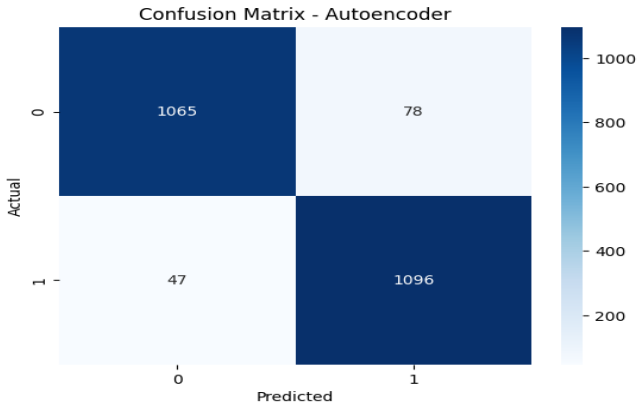


Figure 2 Confusion Matrix of Autoencoders

Autoencoder model also performed very well in phishing detection where significant results were achieved in different evaluation metrics. It attained a probability of 94.53 percent, which means that it made most of the phishing and legal websites correctly. The model is efficient as the accuracy of the model is indicated by an accuracy score of 93.36%. The Autoencoder was found to identify a significant rate of phishing sites, including those that could be more subtle and excellent, and this was achieved at 95.89 with a recall. The F1 score of 94.61 percent further highlights the weighted score on the accuracy and recall such that despite the various forms of phishing attacks, the model works well. The Autoencoder also scored AUC on 0.9876, indicating a great potential of effective results in differenceiating between a phishing and a legitimate site. These findings indicate that the Autoencoder model is very suitable in detecting phishing especially in the sense that it was able to identify unusual and outliers in the data, which are typical of phishing attacks.

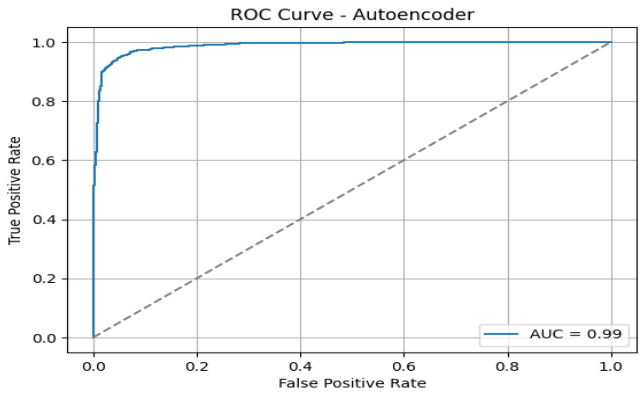


Figure 3 ROC curve of Autoencoder

Feedforward Neural Networks:

Figure 4 Confusion Matrix of DNN

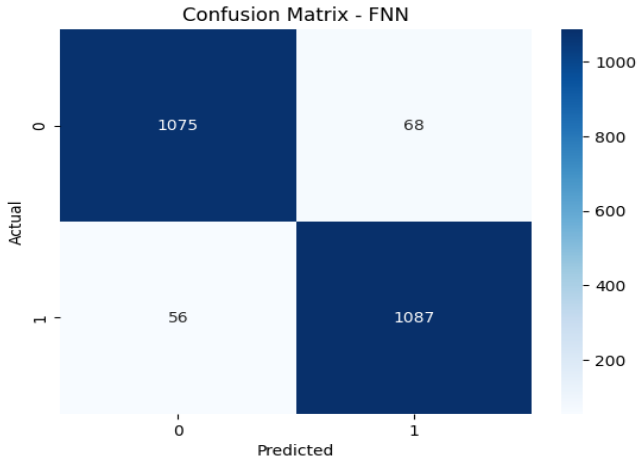


Figure 5 Confusion Matrix of FNN

The Feedforward Neural Network (FNN) model featured good results in phishing detection based on its evaluation measures. The model recorded a rate of 94.58 with accuracy thus depicting a high rate of accurate classification of phishing and legitimate websites. The FNN had a precision of 94.11 meaning that it reduced false positives at a significant level and therefore most of the identified phishing websites are actually malicious. The recall score of 95.10% indicates that the model can identify most phishing attacks, including those ones that can be more explicit or intricate.

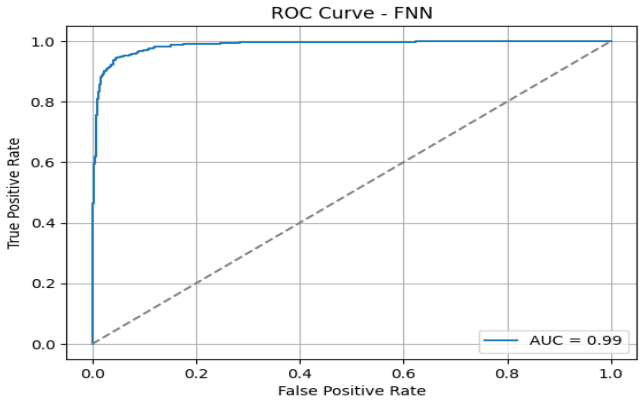
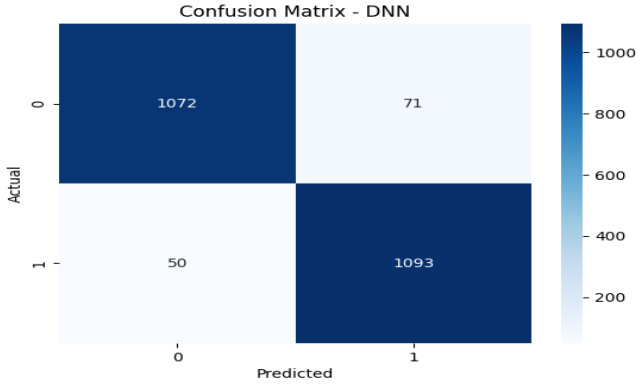


Figure 6 ROC Curve of FNN

There was also a solid overall performance of an F1 score of 94.60. Moreover, FNN attained an area under the curve (AUC) of 0.9871, which was very high in terms of separating phishing and valid websites. These findings suggest that the FNN model is an efficient and useful model of detecting phishing and can be used to detect the malicious websites with high degree of accuracy and reduced errors.

Deep Neural Networks:



The Deep Neural Network (DNN) model analysis has been assessed on the premises of several crucial measures, that is, accuracy, precision, recall, F1 score, and AUC. The DNN had a high accuracy of 94.71%,

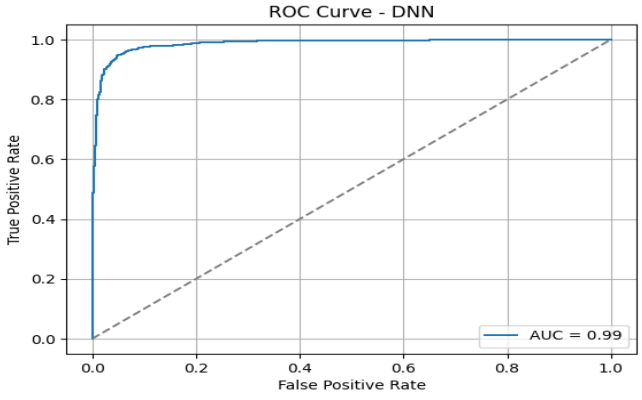


Figure 7 ROC curve of DNN

Precision, which measures the proportion of correctly hit rate on all the websites that were predicted to be phishing, was 93.90, which compared well with minimizing false positives. The model was able to identify phishing attempts by a recall of 95.63 which is excellent as this metric indicates the capacity of the model to identify all the real phishing sites. The F1 score which is the balanced score of precision and recall, is 94.76, showing the overall good performance. The model also had an Area Under the Curve (AUC) of 0.9865 indicating it has outstanding discriminating power that is phishing and legit websites. These findings demonstrate the sound capability of the DNN to identify the presence of phishing attacks and it exhibits the decent balance of precision and recall that makes it a credible source of identifying phishing.

VI. DISCUSSION

The proposed phishing detector that uses both machine learning and deep learning solutions has been found to demonstrate positive results in detecting phishing attacks of various attack vectors. Our paper was focused on enhancing the detection accuracy by means of combining models such as GCN, Autoencoders, FNN, and DNN and the optimal strategy of feature selection techniques. The system can dissect and detect the complex patterns and connections between URLs and domains and email messages through these models that attackers commit numerous hit and miss attacks to defraud users in most cases.

The lessons, which we have gained during our study are that the multifaceted attack of phishing should be addressed with the aid of different models. Though all of the described models have their merits, a combination of GCNs and DNNs is also an effective approach, which can be used to learn how the data is not reflected in a linear, but the structural relationships (between the phishing websites that are interconnected with one another) nevertheless. The ability to learn graph-based dependencies means that GCNs particularly come in handy in cases of phishing in which two or more participants are interacting e.g. phishing email that activates malicious websites. DNNs on the other hand are proficient at learning intricate details with extensive data and it can detect even faint pattern that could lead to phishing.

The other valuable conclusion of our research is the importance of feature selection in the improvement of

the model performance. We just had to be correct and less sophisticated in our calculations through consideration of the most helpful aspects and that is URL structure, domain reputation, and text content in emails. The feature selection process was used to filter out the unnecessary or excessive data that is not necessary and is more target oriented. In addition, the overfitting nature of the complex models in case the models are applied with an excessive amount of features but without the employment of the suitable selection methods was also addressed, and the utilization of autoencoders into the framework was also successful in the context of the problem of anomaly detection. Autoencoders, via reconstruction errors, could classify outliers, or potential phishing attacks, as the normal behavior of authentic websites and emails. This unmonitored system goes hand in hand with the supervised learning systems and it adds another level of strength to the detection system. This is also what makes the model a bit more flexible with the system not expanding less efficient due to the changes in the approaches used by attackers since autoencoders may detect the phishing tactics never witnessed before.

VII. CONCLUSION

We have introduced a developed phishing detecting framework in this study combining machine learning and deep learning in detecting phishing attacks more precisely and more resiliently. Our solution merges models, e.g., GCN, Autoencoders, Feedforward Neural Network (FNN) as well as DNN, and has integrated optimal feature selection to offer a unified solution to the nature of the problem of phishing. The system was having a varied population of legitimate and phishing websites and emails which helped it to learn and identify a variety of phishing activities.

Our experiment trials showed that the hybrid model was more effective than the traditional approaches in phishing detection. The relationship was aptly captured with GCNs between phishing actors, including websites and emails, and insights were given to the structural relationship that can form a manifestation of ill motives. In the meantime, DNNs and FNNs outperformed in recognizing complex, non-linear trends in the data and served to identify phishing attacks based on a new trick or subtle techniques. Autoencoders, conversely, had their part since they were used to mark any anomalous behavior that was captured which provides an extra level of detection to support the supervised learning models.Autoencoders also help increase the efficiency and performance of the system. The choice of the most pertinent features allowed us to simplify the computation and enhance the precision of the model: URL structure, reputation of the domain, and content of the emails. The deployment through a web interface design allowed the system to be available to both technical and non-technical end users such that the phishing detection process was easily deployable on real time platforms.

There are areas that can be improved even though it has been successful. It can be also improved by constantly updating the system with an updated model that will respond to the new phishing tactics. Also, although accuracy was also better in ensemble approach, it also led to complex computations, which in future may be optimized in another version of the model. The detection of the system could be further enhanced by increasing the range of phishing categories to be paid attention to and enhancing the efficiency of the models used to cut down on the processing time.

VIII. FUTURE ENHANCEMENT

Although promising results were observed in the proposed phishing detecting framework, some major aspects require subsequent improvements which would improve the performance and applicability of the framework. The capability to continually refresh and retrain the model is also among the most critical ones. Phishing scams are in the ongoing development, and they are always able to come up with new strategies that will get them past current security measures. To follow these developments, mechanisms of continuous data collection and updated models should be incorporated in the framework. The system will be able to be effective against emerging threats, by ignoring the need to manually add new phishing data to the system and retrain the model.

The other direction of improvement in the future is to increase the size of the dataset to include a more significant part of the phishing methods. The present dataset, despite its heterogeneity, might not be the most recent phishing techniques, perhaps those that apply a new way of social engineering or another sophisticated course of attack such as voice phishing (vishing). The system can be trained to identify a broader range of attack examples, which enhances its capacity to detect new threat types, by broadening the variety of examples of phishing.

Moreover, even though the ensemble method applied under the framework facilitates better level of detection, it also comes with increased computational costs. Further development of the actual work in the future may be aimed at simplifying the model, without affecting its accuracy since it is required to be optimized to allow real-time performance. The computational load could be reduced by techniques like model pruning, quantization, or lighter architectures and help make the system more efficient and faster, particularly when there are other resource-constrained environments.

IX REFERENCES

[1] I. Tyagi, J. Shad, S. Sharma, S. Gaur, and G. Kaur, "A Novel Machine Learning Approach to Detect Phishing Websites," *2018 5th International Conference on Signal Processing and Integrated Networks, SPIN 2018*, pp. 425–430, Sep. 2018, doi: 10.1109/SPIN.2018.8474040.

[2] M. Chawla and S. S. Chouhan, "A Survey of Phishing Attack Techniques," *Int. J. Comput. Appl.*, vol. 93, no. 3, pp. 32–35, May 2014, doi: 10.5120/16197-5460.

[3] A. V. Kumar, A. Prathiba, A. Ashritha, N. H. Reddy, and X. S. A. Shiny, "Phishing Website Detection Based on URL Features," *International Journal of Scientific Research in Engineering & Technology*, vol. 53, no. 02, pp. 73–78, doi: 10.59256/ijsreat.20250502011.

[4] S. S. Ravindra, S. J. Sanjay, S. N. A. Gulzar, and K. Pallavi, "Phishing Website Detection Based on URL," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, pp. 589–594, Jun. 2021, doi: 10.32628/cseit2173124.

[5] C. Opara, Y. Chen, and B. Wei, "Look before you leap: Detecting phishing web pages by exploiting raw URL and HTML characteristics," *Expert Syst. Appl.*, vol. 236, no. 6, p. 121183, Feb. 2024, doi: 10.1016/j.eswa.2023.121183.

[6] R. Gaffer and M. Helali, "Phishing Detection Using Hybrid Machine learning Techniques," vol. 45, pp. 45–52, doi: 10.1654/zkdx.2024.29.2-4.

[7] D. M. Divakaran and A. Oest, "Phishing Detection Leveraging Machine Learning and Deep Learning: A Review," *IEEE Secur. Priv.*, vol. 20, no. 5, pp. 86–95, 2022, doi: 10.1109/MSEC.2022.3175225.

[8] S. Sami, M. Aldaham, O. Ouda, and A. A. Abd El-Aziz, "Improved Detection of Phishing Websites using Machine Learning," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 21s, pp. 4619–4633, Mar. 2024, Accessed: Feb. 28, 2026. [Online]. Available: https://www.ijisae.org/index.php/IJISAE/article/view/6351

[9] N. Puri, P. Saggar, A. Kaur, and P. Garg, "Application of ensemble Machine Learning models for phishing detection on web networks," *2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT)*, pp. 296–303, 2022, doi: 10.1109/CCiCT56684.2022.00062.

[10] N. Puri, P. Saggar, A. Kaur, and P. Garg, "Application of ensemble Machine Learning models for phishing detection on web networks," *Proceedings - 2022 5th International Conference on Computational Intelligence and Communication Technologies, CCICT 2022*, pp. 296–303, 2022, doi: 10.1109/CCiCT56684.2022.00062.

[11] A. K. V, A. A, B. Jose, K. Anilkumar, and O. T. Lee, "Phishing Detection using Machine Learning based URL Analysis: A Survey," *International Journal of Engineering Research & Technology*, vol. 9, no. 13, Aug. 2021, doi: 10.17577/IJERTCONV9IS13033.

[12] Q. E. ul Haq, M. H. Faheem, and I. Ahmad, "Detecting Phishing URLs Based on a Deep Learning Approach to Prevent Cyber-Attacks," *Applied Sciences 2024, Vol. 14, Page 10086*, vol. 14, no. 22, p. 10086, Nov. 2024, doi: 10.3390/app142210086.

[13] J. L. Wilk-Jakubowski, L. Pawlik, G. Wilk-Jakubowski, and A. Sikora, "Machine Learning and Neural Networks for Phishing Detection: A Systematic Review (2017–2024)," *Electronics (Switzerland)*, vol. 14, no. 18, p.

3744, Sep. 2025, doi: 10.3390/electronics14183744.

[14] E. Kytidou, T. Tsikriki, G. Drosatos, and K. Rantos, "Machine learning techniques for phishing detection: A review of methods, challenges, and future directions," *Intelligent Decision Technologies*, vol. 19, no. 6, pp. 4356–4379, Nov. 2025, doi: 10.1177/18724981251366763.

[15] G. Mohamed, J. Visumathi, M. Mahdal, J. Anand, and M. Elangovan, "An Effective and Secure Mechanism for Phishing Attacks Using a Machine Learning Approach," *Processes*, vol. 10, no. 7, Jul. 2022, doi: 10.3390/pr10071356.

[16] S. K. Siva Rama Krishna, "Phishing Website Detection using Machine Learning and Deep Learning Techniques," *Power System Technology*, vol. 49, no. 1, pp. 947–959, Mar. 2025, doi: 10.52783/pst.1643.

[17] U. Zara, K. Ayyub, H. Ullah Khan, A. Daud, T. Alsahfi, and S. Gulzar Ahmad, "Phishing Website Detection Using Deep Learning Models," *IEEE Access*, vol. 12, pp. 167072–167087, 2024, doi: 10.1109/ACCESS.2024.3486462.