A Survey on Phishing Website Detection Using Deep Neural Networks

Vivek Sharma[1(✉)] and Tzipora Halevi[2(✉)]

[1] Department of Computer Science, The Graduate Center, CUNY, New York, USA
vsharma@gradcenter.cuny.edu
[2] Department of Computer Science, Brooklyn College, CUNY, New York, USA
halevi@sci.brooklyn.cuny.edu

Abstract. Phishing is a social engineering attack, where an attacker poses as a legitimate individual or institution and convinces a victim to divulge their details through human interaction. There has been a steep rise in phishing cases across the globe. A report by Cisco [1] shows that phishing was the reason for 90% of data breaches in 2021. Various detection models have been proposed in the past to counter such attacks. Some proposed models work on improving the detection rate of phishing URLs while others focus on reducing their detection time. Authors have used machine learning, deep learning, and various other novel mechanisms in feature selections that result in high algorithm performance. This study is a systematic analysis of recent work utilizing deep learning for phishing detection, highlighting the research methods, algorithms, programming tools, and datasets used in such studies. This study further proposes some guidelines for future research, which include standardizing documentation and performance reporting. These guidelines may help researchers in their quest to replicate others' work and compare newly proposed methods with previously developed systems.

Keywords: Website phishing · Neural network · Survey · Phishing detection

1 Introduction

Phishing attacks continue to be very common with 465 brands targeted in Mar 2021 (Statista [2]). According to [3], phishing incidents rose 220% during the pandemic compared to the otherwise yearly average, with 52% of these attacks targeting brand names. 72% of the attacks during the pandemic used a valid HTTPS certificate while almost all of them used TLS encryption. Phishing website detection can help in finding such attempts and keep everyone safe in today's digital world. A lot of work has been done in the past and this paper analyzes articles related to phishing website detection using deep learning. The motivation of this article is to address the lack of standardization and difficulty in comparing various methodologies in this field. It aims to familiarise its reader

C. Stephanidis et al. (Eds.): HCII 2022, CCIS 1655, pp. 684–694, 2022.
https://doi.org/10.1007/978-3-031-19682-9_87

with the methodologies, algorithms, and tools used in such studies. It also provides statistical figures to summarize the results and adds suggestions that might encourage easy replication and comparison across similar studies.

Research Question: This paper looks at the following research questions: What is the current state of the research in phishing website detection using Deep Learning and how can proposed methodologies be made easier to replicate and comparable with other studies?

2 Theoretical Background

Phishing is a cybercrime where an attacker poses as a legitimate institution to lure the target into providing their sensitive data (Phishing.org [4]). Apart from email and website phishing attempts, there are other variants of such types of attacks that use voice calling and text messages. Additional attacks include website forgery, malware, and domain spoofing through which a victim can be trapped in phishing. Typically, Phishing is used against a large number of random targets, while in spear phishing, a targeted version of phishing, the attacks are targeted towards certain individuals who may possess valuable information.

To counter such attacks, various detection models have been proposed. Some of those concentrate on improving the detection rate of phishing URLs while others focus on reducing the detection time. To achieve this, authors use machine learning models like Naive Bayes, Logistic Regression, Random Forest, Decision Trees, Support Vector Machine (SVM), k-Nearest Neighbour, and deep learning models. Deep learning models include Convolutional Neural Network (CNN), Long Short Term Memory (LSTM), Recurrent Neural Network (RNN), and various other novel mechanisms.

Section 3 in this paper discusses the search strategies, inclusion, and exclusion criteria used in this systematic review. Section 4 covers the various algorithms used, paper goals, contents of the datasets, programming tools, and metrics used in these studies. Section 5 concludes the article with proposed suggestions for standardization techniques for future phishing detection research.

3 Methodology

This search was performed on City University of New York's (CUNY) online library OneSearch. We kept the search criteria broad by searching articles related to phishing and manually narrowed it down to website phishing. This is to ensure that papers that do not have the exact keywords can still be included in the study. Systematic review methodology by Kitchenham [5] is utilized in this study.

3.1 Search Strategy

– **Keyword Search:** The following search strings were used to find relevant papers: "Phishing" AND "Detection"

- **Period:** Articles published between 2017 and 2021
- **Paper Type:** Articles published in conferences or journals
- **Search Database:** City University of New York's online library CUNY One-Search
- **Inclusion Criteria:**
  - Articles written in English.
  - Article scheduled to be published with a pre-print available
  - Articles including keywords in title, full-text or their metadata.
  - Articles implementing or proposing a solution relevant to phishing detection
- **Exclusion Criteria:**
  - Book Chapters, Newsletter Articles, Books, and Dissertation.
  - Systematic reviews and literature survey.
  - Articles with pure ML-based implementations.

The search resulted in ninety-two papers out of which twenty-six papers were discarded after applying the exclusion criteria leaving a total of sixty-six papers. Articles that utilized pure Machine learning-based implementation such as kNN, SVM, Random Forest, and Logistic Regression were excluded from this study, except for hybrid and ensemble models where some ML algorithms are combined with deep learning-based algorithms.

4 Discussion/Findings Overview

Once the papers were collected and filtered using the method specified in the previous section, the articles were examined and categorized according to different criteria, including:

- **Datasets:** Datasets are used in the training and testing of the model. In phishing detection, the data needs to be continuously updated so researchers list out the methodologies used to fetch data from popular data sources. The Datasets include different features such as URLs, length of URL, domain based-features including the age of domain, DNS record, and HTML based features: number of out links, anchor tags, etc. Table 1 lists the datasets and data sources that are shared and used in multiple studies.
- **Programming tools:** Listing out the programming tools helps researchers in reproducing the work and comparing their proposed work against the same environmental specifications. The result in Sect. 4.2 confirms the recent trend in the use of python over other programming languages.
- **Algorithms used:** Various. deep learning based algorithms are used in model training. Some researchers fused multiple algorithms in ensemble and hybrid approaches to improve detection accuracy of the model.
- **Research Methods:** The different design goals of the covered work are described, which include removing dependencies or minimizing the needed input data as well as improvement of detection rates and reducing training and testing runtime.

4.1 Datasets

While some studies use proprietary datasets, multiple studies include publicly accessible ones, listed in Table 1.

Table 1. Most popular datasets used in phishing detection

Source	Details	Continuously updated
PhishTank [6]	Phishing URL are submitted and updated by registered users of its community. Users can fetch data through API key	✓
Common crawl [7]	Web crawled data which can be accessed through HTTP or S3. Column like IP address, URL hostname, port, protocol, query, URL hostname, and target URL to name few	✓
Alexa [8]	Top sites are listed based on their traffic ranks which is computed based on average daily visitor and page views	✓
DMOZ [9]	It was earlier known as Open Directory Project (ODP). RDF dumps of database are available to download from the site	✓
Phishload [10]	Contains more than 1000 targeted legitimate websites. Dataset contains HTML source code and other information like id, alexa rank, URL, URL has etc. The size of this dataset is roughly 6GB with screenshots and without screenshots is 44MB.	✗
UCI Phishing dataset [11]	Contains 2456 instances with 30 attributes. Contains IP address, URL-based features, HTML-based features and domain-based features like domain age, DNS record, Page rank etc. Training data is in .arff format and the size is less than a MB.	✗
Kaggle [12]	Contains 1353 instances with 10 attributes each. The features are URL, URL length, IP address, prefix/suffix, domain age etc. Phishing websites were selected from Phishtank and legitimate website were extracted from Yahoo. There are 548 legitimate, 103 suspicious, and 702 phishing websites labelled as 1, 0, −1 respectively	✗
PhishStorm [13]	Contains 48,009 legitimate URL and same number of phishing URL, taking the total to 96,018 URLs. Data is in .csv format and is approximately 3MB in size. This dataset is described and first used in study by Marchal et al. [14]	✗
Openphish [15]	Contains attributes like hostname, URL, path, SSL metadata, IP, targetted brand etc. Datasets dumps ranges from 30–180 days of phishing data. Provides an SQLite dataset which can be easily integrated using an open-source API [16]. Screenshot are available for most of the URLs.	✓

Datasets Features: Different types of data are used in various research, including:

– URL: Uniform Resource Locator(URL) and its related information like its length, and use of special characters with or without trimming were used in most of the studies. While most studies use URL repositories, some combine it with additional data listed below

- Metadata information: Metadata website information includes age of domain, popularity of websites, DNS rank, etc.
- Webpage content: It includes HTML tags based on information like the number of links in the source code.
- Images/screenshots: Some datasets have images and screenshots of the website or logo of targeted brands. These images were used along with URL based information to improve the accuracy of the model.

4.2 Programming Tools

Few articles described the programming tools and specifications used in their experiments. Among them almost 35% of the articles used python and approximately 10 % of them used WEKA, Java, and MATLAB-based implementation. The use of these latter languages is decreasing as compared to python. A survey conducted by Kaggle [17] indicated there is a large number of submissions in python and a significantly lower rate of submissions in other languages such as MATLAB and Java. This trend has been observed starting 2013 [18]

4.3 Algorithms

This section lists major deep learning algorithms used in training the classification models. The literature shows three main approaches used by researchers: deep learning models with a single algorithm, hybrid approaches, and ensemble approaches. The latter two approaches are analyzed in more detail in Sect. 4.4

Deep Learning Models: Primary deep learning models used were Deep Neural Networks with hidden layers, CNN, LSTM, and RNN. The models and their accuracy using these algorithms are shown in Table 3.

Ensemble Models: Multiple diverse models are generated and a final prediction is made after aggregating their predictions. Although this model consists of several base models, the model still acts and works like a single model. The ensemble model aims to reduce the generalization error of prediction. Nagaraj et al. [19] used random forest and neural network to get an accuracy of 93.41 on their ensemble model. Another model with LSTM and SVM saw an accuracy in the range of 95.40%–98.50%

Hybrid Models: The model is made by fusing multiple models into a single model. The algorithms used in such models and accuracy ranges are presented in Table 2.

4.4 Research Methods

This section categorizes the research methods used in the surveyed studies. We broadly classify the domain where these improvements were visible into three categories.

- **Novelty in Feature Selection techniques:** [20–23] used various feature extraction methods. [24,25] introduced novel features and [26] evaluated its model on different feature spaces. [27,28] used novel feature selection technique like Recursive Feature Elimination(RFE) in their work.
- **Use of Fusion/multilevel architecture:** Study by Kazienko et al. [29] shows the use of fusion and multilevel techniques like ensemble model and hybrid model improves the performance of the ML model. While ensemble models can take more time to train the model, a study by Sameen et al. [30] speeds this up through the use of a multi-threaded approach. Different fusion/multilevel architecture models are presented in Table 2.
- **Generation of Adversarial URLs:** Evaluating security aspects of a model is useful for evaluating the ability to prevent adversarial attacks. [31] assesses vulnerability of a system while [32] talks about defense against attacks. Adversarial phishing URL were generated by [33–37].
- **Eliminating need of dependencies:** Performance of a model can be affected by the interruption of third-party services, language dependencies, etc. Study by Somesha et al. [38], Yang et al. [39], Waziral et al. [27], and Jain et al. [40] eliminated need for third-party services in their work. Webpage content-based features, language dependencies and use manually crafted features were eliminated in [39,41,42] respectively.
- **Additional methods:** Work by [22,43–45] were directed towards increasing speed of detection. [46,47] addressed zero-day phishing vulnerabilities, [48] visualized internal working of a DNN while [41,49,50] created phishing detection aimed at low-power mobile devices.

Table 2. Hybrid approaches used in studies

Models	Author	Accuracy (%)
DNN - BiLSTM	Ozcan et al. [20]	99.21
DBN - SVM	Yu et al. [51]	99.96
AE - CNN	Zhang et al. [52]	97.68
CNN - SVM	Zhang et al. [52]	97.68
CNN-LSTM	Adebowale et al. [53]	92.10–93.28
	Yang et al. [54]	98.99
	Bu et al. [46]	95.40–98.32
CNN - RF	Yang et al. [39]	99.25–99.35
CNN - BiLSTM	Feng et al. [55]	99.05
	Zhang et al. [43]	98.03–99.79
	Zhang et al. [23]	92.09–98.84

Table 3. Deep learning algorithms used in studies

Models	Author	Accuracy (%)
DNN	Sumathi et al. [56]	90
	Lakshmi et al. [57]	92.09–98.44
	Somesha et al. [38]	99.43
	Soon et al. [58]	94.27–94.41
LSTM	Somesha et al. [38]	99.57
	Hashim et al. [32]	98.65
	Su et al. [59]	99.1
	Desuoza et al. [60]	95.89–98.30
	Pham et al. [36]	97
CNN	Wei et al. [49]	83.57–86.63
	Somesha et al. [38]	99.52
	Bartoli et al. [61]	98.2–99.2
	Al-Alyan et al. [62]	88.54–98.22
	Singh et al. [63]	98
	Mourtaj et al. [21]	97.94
	Aljofey et al. [64]	51.29–98.58
	Korkmaz et al. [22]	88.90
	Yerima et al. [65]	95.80–98.20
	Jawade et al. [45]	99
RNN	Feng et al. [48]	99.05
	Dutta et al. [66]	98.03–99.79
	Bahnsen et al. [67]	92.09–98.84

4.5 Recommendations

- **Use of shared datasets:** Shared datasets can help in replication and comparison among different models. It would be useful for researchers to test their data on shared datasets in addition to any proprietary dataset when possible, to help improve the side-by-side evaluation of different algorithms. Sometimes the datasets are not shared due to privacy or ethical issues. The researchers can share their approach to fetching data from a data source for easier replication.
- **Sharing code/algorithm:** This will encourage reproducibility of the work and provide a way for researchers to further adapt or expand the current work.
- **Testing on updated datasets continuously.** As new phishing websites are continuously introduced, this will provide a way to gauge the performance of successful phishing detection algorithms on newly introduced phishing URL
- **Standardize testing environment documentation:** Experimentation is the description of the environment in which the experiment was performed.

Creating a standard method of documentation, which will include details regarding the system parameters used and run-time, can help researchers assess the usability of different methods in different attack scenarios as well as recreate the test environment in future studies.

5 Conclusion

Phishing can be done through different techniques. This work focuses on the detection of phishing websites using deep learning neural networks. This study found that there is a growing body of research in this field, utilizing different techniques, datasets, and attack scenarios. This work points to share as well as datasets that continue to update and can be used in future research. It also compares the goals and design details of different studies and the resulting reported performance. This paper suggests methods for standardization of algorithms and testing reports, which can help improve the design of future studies.

References

1. CISCO: cisco threat report 2021. https://umbrella.cisco.com/info/2021-cyber-security-threat-trends-phishing-crypto-top-the-list
2. Johnson, J.: Phishing - statistics & facts. https://www.statista.com/topics/8385/phishing/
3. labs, F.: Phishing attacks soar 220% during COVID-19 peak as cybercriminal opportunism intensifies. https://www.f5.com/company/news/features/phishing-attacks-soar-220-during-covid-19-peak-as-cybercriminal
4. phishing.org: what is phishing. https://www.phishing.org/what-is-phishing
5. Kitchenham, B.: Procedures for performing systematic reviews. Keele, UK, Keele University **33**(2004), 1–26 (2004)
6. PhishTank: PhishTank. https://phishtank.org/
7. Crawl: common crawl. https://commoncrawl.org/
8. Alexa: alexa top sites. https://www.alexa.com/topsites
9. DMOZ: Dmoz phishing dataset. https://dmoz-odp.org/docs/en/rdf.html
10. Maurer, M.: Phishload. https://www.medien.ifi.lmu.de/team/max.maurer/files/phishload/index.html
11. UCI: UCI phishing dataset. https://archive.ics.uci.edu/ml/datasets/phishing+websites
12. Kaggle: kaggle. https://www.kaggle.com/ahmednour/website-phishing-data-set
13. Marchal, S: PhishStorm. https://research.aalto.fi/en/datasets/phishstorm-phishing-legitimate-url-dataset
14. Marchal, S., François, J., State, R., Engel, T.: PhishStorm: detecting phishing with streaming analytics. IEEE Trans. Netw. Serv. Manage. **11**(4), 458–471 (2014)
15. OpenPhish: OpenPhish. https://openphish.com/phishing_database.html
16. OpenPhish: OpenPhish API. https://github.com/openphish/pyopdb
17. Kaggle: Kaggle survey 2019. https://www.kaggle.com/kaggle-survey-2019
18. Brownlee, J.: Best programming language. https://machinelearningmastery.com/best-programming-language-for-machine-learning/

19. Nagaraj, K., Bhattacharjee, B., Sridhar, A., Sharvani, G.: Detection of phishing websites using a novel twofold ensemble model. J. Sys. Inf. Technol. (2018)
20. Ozcan, A., Catal, C., Donmez, E., Senturk, B.: A hybrid DNN-LSTM model for detecting phishing URLs. Neural Comput. Appl. 1–17 (2021)
21. Mourtaji, Y., Bouhorma, M., Alghazzawi, D., Aldabbagh, G., Alghamdi, A.: Hybrid rule-based solution for phishing URL detection using convolutional neural network. Wirel. Commun. Mobile Comput. **2021** (2021)
22. Korkmaz, M., Kocyigit, E., Sahingoz, O.K., Diri, B.: Phishing web page detection using N-gram features extracted from URLs. In: 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), pp. 1–6. IEEE (2021)
23. Zhang, Q., Bu, Y., Chen, B., Zhang, S., Lu, X.: Research on phishing webpage detection technology based on CNN-BiLSTM algorithm. In: Journal of Physics: Conference Series, vol. 1738, p. 012131. IOP Publishing (2021)
24. Yi, P., Guan, Y., Zou, F., Yao, Y., Wang, W., Zhu, T.: Web phishing detection using a deep learning framework. Wirel. Commun. Mobile Comput. **2018** (2018)
25. Xiao, X., Zhang, D., Hu, G., Jiang, Y., Xia, S.: CNN-MHSA: a convolutional neural network and multi-head self-attention combined approach for detecting phishing websites. Neural Netw. **125**, 303–312 (2020)
26. Liu, D.J., Geng, G.G., Jin, X.B., Wang, W.: An efficient multistage phishing website detection model based on the case feature framework: aiming at the real web environment. Comput. Secur. **110**, 102421 (2021)
27. Wazirali, R., Ahmad, R., Abu-Ein, A.A.K.: Sustaining accurate detection of phishing URLs using SDN and feature selection approaches. Comput. Netw. **201**, 108591 (2021)
28. Saha, I., Sarma, D., Chakma, R.J., Alam, M.N., Sultana, A., Hossain, S.: Phishing attacks detection using deep learning approach. In: 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 1180–1185. IEEE (2020)
29. Kazienko, P., Lughofer, E., Trawinski, B.: Editorial on the special issue "hybrid and ensemble techniques in soft computing: recent advances and emerging trends". Soft. Comput. **19**(12), 3353–3355 (2015). https://doi.org/10.1007/s00500-015-1916-x
30. Sameen, M., Han, K., Hwang, S.O.: PhishHaven-an efficient real-time AI phishing URLs detection system. IEEE Access **8**, 83425–83443 (2020)
31. Ogawa, Y., Kimura, T., Cheng, J.: Vulnerability assessment for deep learning based phishing detection system. In: 2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), pp. 1–2. IEEE (2021)
32. Hashim, A., Medani, R., Attia, T.A.: Defences against web application attacks and detecting phishing links using machine learning. In: 2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), pp. 1–6. IEEE (2020)
33. AlEroud, A., Karabatis, G.: Bypassing detection of URL-based phishing attacks using generative adversarial deep neural networks. In: Proceedings of the Sixth International Workshop on Security and Privacy Analytics, pp. 53–60 (2020)
34. Xiao, X., et al.: Phishing websites detection via CNN and multi-head self-attention on imbalanced datasets. Comput. Secur. **108**, 102372 (2021)
35. Zhang, J., Li, X.: Phishing detection method based on borderline-smote deep belief network. In: Wang, G., Atiquzzaman, M., Yan, Z., Choo, K.-K.R. (eds.) SpaCCS 2017. LNCS, vol. 10658, pp. 45–53. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-72395-2_5

36. Pham, T.D., Pham, T.T.T., Hoang, S.T., Ta, V.C.: Exploring efficiency of GAN-based generated URLs for phishing URL detection. In: 2021 International Conference on Multimedia Analysis and Pattern Recognition (MAPR), pp. 1–6. IEEE (2021)
37. Shirazi, H., Bezawada, B., Ray, I., Anderson, C.: Adversarial sampling attacks against phishing detection. In: Foley, S.N. (ed.) DBSec 2019. LNCS, vol. 11559, pp. 83–101. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-22479-0_5
38. Somesha, M., Pais, A.R., Rao, R.S., Rathour, V.S.: Efficient deep learning techniques for the detection of phishing websites. Sādhanā **45**(1), 1–18 (2020). https://doi.org/10.1007/s12046-020-01392-4
39. Yang, R., Zheng, K., Wu, B., Wu, C., Wang, X.: Phishing website detection based on deep convolutional neural network and random forest ensemble learning. Sensors **21**(24), 8281 (2021)
40. Jain, A.K., Gupta, B.B.: A machine learning based approach for phishing detection using hyperlinks information. J. Ambient. Intell. Humaniz. Comput. **10**(5), 2015–2028 (2019). https://doi.org/10.1007/s12652-018-0798-z
41. Rao, R.S., Vaishnavi, T., Pais, A.R.: PhishDump: a multi-model ensemble based technique for the detection of phishing sites in mobile devices. Pervasive Mob. Comput. **60**, 101084 (2019)
42. Tajaddodianfar, F., Stokes, J.W., Gururajan, A.: Texception: a character/word-level deep learning model for phishing URL detection. In: ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2857–2861. IEEE (2020)
43. Zhang, L., Zhang, P.: PhishTrim: fast and adaptive phishing detection based on deep representation learning. In: 2020 IEEE International Conference on Web Services (ICWS), pp. 176–180. IEEE (2020)
44. Yuan, H., Yang, Z., Chen, X., Li, Y., Liu, W.: URL2vec: URL modeling with character embeddings for fast and accurate phishing website detection. In: 2018 IEEE International Conference on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom), pp. 265–272. IEEE (2018)
45. Jawade, J.V., Ghosh, S.N.: Phishing website detection using fast. ai library. In: 2021 International Conference on Communication information and Computing Technology (ICCICT), pp. 1–5. IEEE (2021)
46. Bu, S.J., Cho, S.B.: Integrating deep learning with first-order logic programmed constraints for zero-day phishing attack detection. In: ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2685–2689. IEEE (2021)
47. Bozkir, A.S., Aydos, M.: LogoSENSE: a companion HOG based logo detection scheme for phishing web page and E-mail brand recognition. Comput. Secur. **95**, 101855 (2020)
48. Feng, T., Yue, C.: Visualizing and interpreting RNN models in URL-based phishing detection. In: Proceedings of the 25th ACM Symposium on Access Control Models and Technologies, pp. 13–24 (2020)
49. Wei, B., et al.: A deep-learning-driven light-weight phishing detection sensor. Sensors **19**, 4258 (2019). https://doi.org/10.3390/s19194258.https://www.mdpi.com/1424-8220/19/19/4258
50. Haynes, K., Shirazi, H., Ray, I.: Lightweight URL-based phishing detection using natural language processing transformers for mobile devices. Procedia Comput. Sci. **191**, 127–134 (2021)

51. Yu, X.: Phishing websites detection based on hybrid model of deep belief network and support vector machine. In: IOP Conference Series: Earth and Environmental Science, vol. 602, p. 012001. IOP Publishing (2020)
52. Zhang, X., Shi, D., Zhang, H., Liu, W., Li, R.: Efficient detection of phishing attacks with hybrid neural networks. In: 2018 IEEE 18th International Conference on Communication Technology (ICCT), pp. 844–848. IEEE (2018)
53. Adebowale, M.A., Lwin, K.T., Hossain, M.A.: Intelligent phishing detection scheme using deep learning algorithms. J. Enterp. Inf. Manage. (2020)
54. Yang, P., Zhao, G., Zeng, P.: Phishing website detection based on multidimensional features driven by deep learning. IEEE access **7**, 15196–15209 (2019)
55. Feng, J., Zou, L., Ye, O., Han, J.: Web2vec: phishing webpage detection method based on multidimensional features driven by deep learning. IEEE Access **8**, 221214–221224 (2020)
56. Sumathi, K., Sujatha, V.: Deep learning based-phishing attack detection. Int. J. Recent Technol. Eng. (IJRTE) **8**(3) (2019)
57. Lakshmi, L., Reddy, M.P., Santhaiah, C., Reddy, U.J.: Smart phishing detection in web pages using supervised deep learning classification and optimization technique ADAM. Wireless Pers. Commun. **118**(4), 3549–3564 (2021). https://doi.org/10.1007/s11277-021-08196-7
58. Soon, G.K., Chiang, L.C., On, C.K., Rusli, N.M., Fun, T.S.: Comparison of ensemble simple feedforward neural network and deep learning neural network on phishing detection. In: Alfred, R., Lim, Y., Haviluddin, H., On, C.K. (eds.) Computational Science and Technology. LNEE, vol. 603, pp. 595–604. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-0058-9_57
59. Su, Y.: Research on website phishing detection based on LSTM RNN. In: 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), vol. 1, pp. 284–288. IEEE (2020)
60. de Souza, C.H.M., Lemos, M.O.O., da Silva, F.S.D., Alves, R.L.S.: On detecting and mitigating phishing attacks through featureless machine learning techniques. Internet Technol. Lett. **3**(1), e135 (2020)
61. Bartoli, A., De Lorenzo, A., Medvet, E., Tarlao, F.: Personalized, browser-based visual phishing detection based on deep learning. In: Zemmari, A., Mosbah, M., Cuppens-Boulahia, N., Cuppens, F. (eds.) CRiSIS 2018. LNCS, vol. 11391, pp. 80–85. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-12143-3_7
62. Al-Alyan, A., Al-Ahmadi, S.: Robust URL phishing detection based on deep learning. KSII Trans. Internet Inf. Syst. (TIIS) **14**(7), 2752–2768 (2020)
63. Singh, S., Singh, M., Pandey, R.: Phishing detection from URLs using deep learning approach. In: 2020 5th International Conference on Computing, Communication and Security (ICCCS), pp. 1–4. IEEE (2020)
64. Aljofey, A., Jiang, Q., Qu, Q., Huang, M., Niyigena, J.P.: An effective phishing detection model based on character level convolutional neural network from URL. Electronics **9**(9), 1514 (2020)
65. Yerima, S.Y., Alzaylaee, M.K.: High accuracy phishing detection based on convolutional neural networks. In: 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS), pp. 1–6. IEEE (2020)
66. Dutta, A.K.: Detecting phishing websites using machine learning technique. PLoS ONE **16**(10), e0258361 (2021)
67. Bahnsen, A.C., Bohorquez, E.C., Villegas, S., Vargas, J., González, F.A.: Classifying phishing URLs using recurrent neural networks. In: 2017 APWG Symposium on Electronic Crime Research (eCrime), pp. 1–8. IEEE (2017)