Detection of E-Mail Phishing Attacks – using Machine Learning and Deep Learning

Dhruv Rathee
Student
Maharaja Surajmal Institute of Technology,
GGSIPU, New Delhi

Suman Mann
Associate Professor
Maharaja Surajmal Institute of Technology,
GGSIPU, New Delhi

ABSTRACT
Phishing is the most prominent cyber-crime that uses camouflaged e-mail as a weapon. In simple words, it is defined as the strategy adopted by fraudsters in-order-to get private details from persons by professing to be from well-known channels like offices, bank, or a government organization. In this era of modernization, electronic mails are accustomed globally as communiqué channel for both private and professional purposes. The particulars exchanged over e-mails are often confidential and sensitive for example info of bank statements, payment bills, debit-credit reports, and authentication data. This makes e-mails precious for hackers because they can exploit these details for maleficent intends. The main goal of the attackers is to acquire personal details by deceiving the e-mail recipient to click noxious link or download the attachment under false pretences. In the last few years, there is an exponential rise in cyber threats including the major ones, phishing e-mails have result in huge monetary and identity losses. Several models have been developed to separate ham and phished e-mails but attackers are always trying new methods to invade the privacy of the people. Hence, there is dire need to perpetually develop new models or to upgrade the existing ones. The focus of the paper is to elaborate that specifically centers around on both machine-learning (ML) and deep-learning (DL) approaches for detecting phishing e-mails. It shows comparative analysis and assessment of various DL and ML models that were proposed in the last few decades to classify phishing e-mails at different stages of crime in a systematic manner. This paper discusses the problem's concept, its explication, and the anticipated future directions.

Keywords
Email Phishing, Phishing, Machine Learning, Deep Learning

1. INTRODUCTION
Phishing e-mail are special kind of spam messages where a criminal also known as an attacker sends victim a fake e-mail that reportedly claims to be originating from legal organizations. These e-mails or messages contain embedded malicious attachments or noxious Uniform Resource Locators (URLs) which installs malware inside the user's system. Malware causes system failure due to the vandalization of internal components of the operating system. Also, the phished e-mail may redirect the user to fraudulent websites that lead to the loss of sensitive information like exploitation of details of banking accounts, login credentials, credit card information, and many more. The main motive of criminals behind this type of scam is monetary gain.

Phishing incidents increased by an astounding amount during the height of worldwide pandemic scares of coronavirus. On average, businesses globally lose $17,700 every minute due to phishing attacks. Incidents involving payment and invoice fraud amplified by 112% between Q1 2020 and Q2 2020, showing phishers were more focused on money, as seen in figure 1, the financial institutions were the top phishing target [2]. The rise of phishing attacks has resulted in significant losses for countries due to thefts of identity and money. Countries are now imposing laws that will prosecute those found guilty. Companies and organizations are also educating their employees about all types of the crime of identity theft. Internet service providers are also not excluded from risk management measures associated with the theft of sensitive information. They have developed various ways to filter and block suspicious e-mails concerning sensitive identity theft. As the crime of theft of sensitive information emerges, people will always find new ways to reduce the risks associated with phishing, the fortuity of stealing sensitive information will equally continue to upsurge. Criminals will always find new ways to deceive people. Therefore, the war will continue until one force uses greater power to overcome the other.

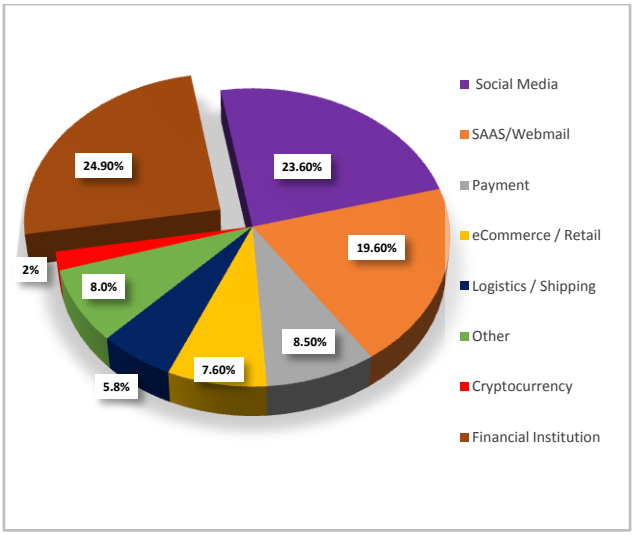

Figure 1- Industries targeted by phishers in first quarter of 2021

Phishing e-mails possess social engineering threat which leads to extraction of personal details from innocent victims [1]. There are various communication channels used for this purpose like SMS, e-mails, voice messages, QR codes, and phone calls, but e-mails are the most common channel used by criminals for phishing attacks [3]. Hence, this survey focuses more on e-mail communiqué in this literature. Identifying phishing messages and e-mails is inevitably a challenging task, as perceived. Since there are enumerable approaches discussed in various papers built using DL and ML to recognise phishing e-mails this survey provides an

organized guide to the previous and current state of the literature. This survey compares and analyses various detection techniques, in addition to identifying and categorizing them. The paper discusses several approaches based on ML classification algorithms and DL methods for detecting phishing e-mails. The paper starts with section 2, it discusses the literature review of anti-phishing strategies also contains a well-structured and organized table of various phishing e-mail detection techniques. Section 3 presents phishing and its lifecycle. Section 4 shows the depiction of phishing e-mails to raise awareness among users that would improve phishing attack detection. Both Section 5 and Section 6 talks over different types of ML and DL approaches to detect phished e-mails. Section 7 discusses survey result. Lastly, section 8 tells about the conclusion and future scope of the paper.

2. LITERATURE REVIEW

In recent years, several works and reviews have been published [29…34], providing crucial knowledge for researchers to understand various approaches to detect phishing. A. Hamid I.R. et al. [7] suggested a mixed-selection model based on the combination of both behaviour and content-based that would help to detect the attacker by using e-mail headers. Aburrous, M. et al. [8] proposed a fuzzy logic-based model by using fuzzy data mining algorithms and their tentative outcomes indicated the prominence of URL and Domain Identity in detection of website phishing. In the study [9], Varshney scrutinizes, evaluates, and distinguishes majorly all significant and novel models discovered in the branch of fraudulent website detection. Recently, Vijayalakshmi M. discusses the past phishing trends, taxonomy and listed state-of-the-art approaches for each category published in the literature review. The authors categorized all solutions for the theft of sensitive information into different categories according to their input parameters such as web-based methods divided into list-based methods, heuristic rules, and learning-based methods. In addition, web-based content solutions were broken down into rule and ML-based solutions. They compared all methods based on segment performance, limitations, external-company service independence, and zero-hour attack detection. Further, the model suggested that the hybrid methods would achieve a higher level of accuracy and suitability for real-time systems. Finally, they concluded that deep learning-based solutions would be an important guide in the future [10]. Also, Said Salloum et al. [11] tells about multiple modern approaches developed using DL and natural-language-processing (NLP) methods of recognizing phishing e-mails along with their limitations and drawbacks.

3. PHISHING AND ITS LIFECYCLE

Phishing is the simplest form of cyber-attack and, simultaneously, the most operational and harmful with an objective of enticing humans to get secluded details like passwords, bank receipts, and account IDs. This is because it attacks the most dangerous and powerful machine on the planet. Phishers are not attempting to utilize the technological weakness in the device's operating system, they're using social engineering. From Windows and iPhones, to Macintosh and Androids, no OS is entirely safe from phishing, regardless of its powerful protection. Infact, attackers frequently go to phishing because they can't see any technological vulnerabilities. This type of cyber-attack is usually triggered by e-mails, instant messages, or phone calls. A flowchart of phishing lifecycle is presented below that discusses the complete process adopted by criminals.

Phishing follows a lifecycle as shown above in figure 2. First, the attacker creates a phishing website that has close correspondence with the official website. For this, criminals use techniques like similar alphabetic characters, spelling errors, and other procedures to build a legitimate website URL, especially domain name and network resource domain. For example, link "https://aimazon.amz-z7acyrojdd0j9i16.xyz/v" mimics https://www.amazon.com. Although, browser on computer can detect a URL address by moving the mouse over a clickable link. It is difficult for the average user to identify these URLs with the naked eye and memory as replica of official URLs. On the other hand, copying of original site's content is also a crucial stage. Often, attackers use scripts to extract web structures, text, and logos of legitimate web pages. Form submission pages like payment page, the password recovery page, and the login page that require the recipient to enter confidential details often deceive by cyber-criminals,.

Secondly, sending the e-mail that purposely misleads victims to click the link, the tactic of sending phished- links is not only by e-mail but also by spoofing mobile applications, quick response (QR) codes, voice messages, and short message service (SMS). With the widespread use of smartphones and social media, the number of channels for criminals to spread false information has increased. In all these processes, images and texts are commonly acquired to deceive recipients into clicking on the link.

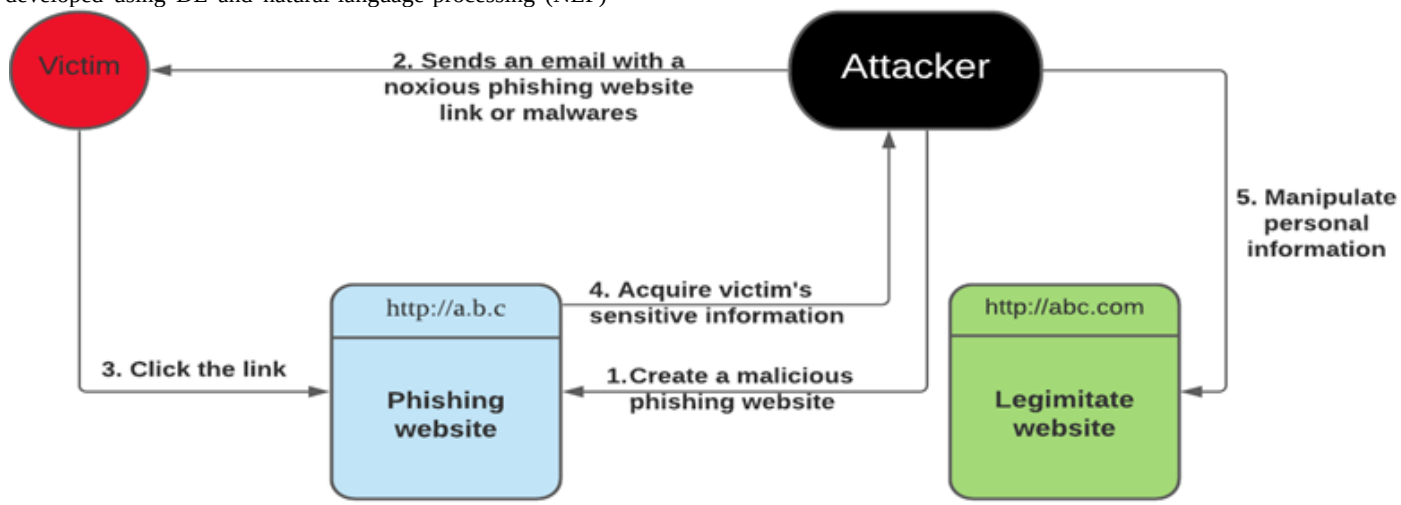

Figure 2- Phishing lifecycle

Even though scam e-mails are sent randomly, every time there is a small proportion of people with weak anti-phishing alertness who would be spoofed.

In this step, the attacker applied social engineering methods, comprising psychological manipulation, to dupe people into making security faults. Perpetrators are experienced at building a sense of urgency and fear and gaining the user's trust via text messages. Afterward, the user clicks the link that will direct them to open a fraudulent website. Particularly, real URL strings are hidden before redirecting to web browsers on mobile phones.

Next step is gathering confidential details on forged site that appears like the real company or corporate web-page using a visual design of the identical logo and content, frequently occurring with payment, reset password, replenished private, and login details. When victims submit these delicate details to web-servers which criminals make, attackers would obtain the desired data of victim.

The final procedure is slinking the recipient's monetary funds via their real details to counterfeit recipient's request for an original webpage. Even some individuals are using identical passwords and usernames for numerous online sites. In this manner, the attacker acquires details of numerous accounts from the victim.

4. PHISHING E-MAILS

Phishing e-mails have transformed and modified into numerous categories, some are well-crafted to appear legitimate such as figure-4, and others like figure-3 are easily recognizable. On the one hand, phished e-mail in figure-4 is specially designed to appear innocuous and has a close resemblance with an e-mail from a legitimate organization, however by clicking on the button, the spoofed website might execute and download some malware or transfer the user to fake websites [5]. On the other hand, the second phished e-mail in figure-3 targets on greediness of human beings, that may not easily fool the person to the fact that by no means they have partook or perceived in this sort of sweepstake, individually winning one million dollars($1,000,000) award money, but greed of human beings for easy funds tends to persuade the victim to open the (.rtf.txt) attachment that often contains a macro, which would kick start the attack Nowadays, some e-mails dupe the users by warning them about the virus attack in their system and provide the solution in the attachment, if they click and install the solution, their system will be trouble-free, and the virus will not attack it.

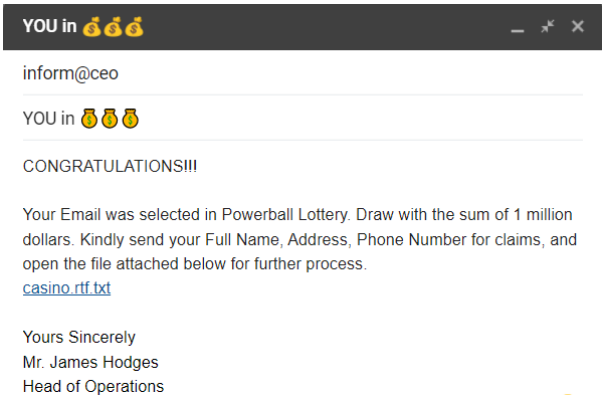

Figure 3 – Easy to spot sample phishing e-mail

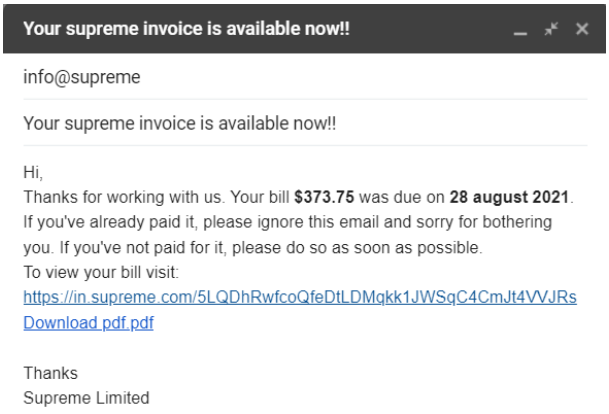

Figure 4 – Phishing e-mail close to legitimate e-mail

The phishing e-mails are sent from authorized accounts, so that the mails seems to appear legit. To accomplish their mission, attackers often keep track of the user personal details. They always send mails from only those accounts that belong to a friend or a previous business colleague by tricking the victim with fake e-mail IDs. The criminals often work very hard to make e-mails appear genuine by entailing believable-wordings, graphic-interface, and logos. Although, phished e-mails do not appear similar as actual phishing e-mails are custom-built for their anticipated objectives. There are a lot of categories and variants of phished e-mails described in the literature [4]. It cannot be helped, but notice the efforts of attackers to design these e-mails and redirection websites appear authentic and benign. The body of the e-mails are carefully crafted to appear trustworthy. Nevertheless, you can observe a lot recognizable or distrustful characteristics for example easy-to-spot sample phishing e-mail attachments, inaccuracy of the content , easy-to-spot sample phishing e-mail attachments, urgent request, the existence of a hyperlink Fig. 3. In certain cases, the source e-mail is absolute doubtful [6]. Regardless of the noticeable characteristics of all the earlier e-mails, approximately most of the average victims with less cyber-awareness are robbed by such scams.

5. PHISHING E-MAIL DETECTION USING MACHINE LEARNING

Machine learning-based approaches help in detecting phishing attacks more efficiently by giving lower false-positive rates and high accuracy in comparison with other methods [12]. Earlier, one of the interesting methods titled PILFERS was proposed by Fette et al. [13] based on 10 features that mostly examine URL and presence of JavaScript to flag e-mails as phished. Nine features were extracted from the e-mail and the last few features were obtained from the WHOIS query. They used larger datasets of about 7000 normal e-mails and 860 phishing e-mails for training and testing the classifier. They focused specifically on URL properties which might not be the appropriate technique because identification of phished e-mails depends on various factors. Also, criminals could use tools to obfuscate URLs such as tiny URLs (https://tiny.qe/) and design them to appear legal. Their filter scores 97.6% F-measure, false-positive rate of 0.13%, and a false-negative rate of 3.6% respectively. Abu-Nimeh et al. [14] study the performance of different classifiers used in text mining such as Support Vector Machines (SVM), logistic regression (LR), Bayesian additive regression trees (BART), classification and regression trees (CART), neural networks(NN), and random forests (RF). They test on a publically available dataset collection of about 1700 phishing mails and 1700 legitimate mails from private mailboxes. The training and testing

included 10-fold cross-validations, and it was found that RF produces the best result with F-Measure 90.24%, but at the same time, it has the highest false-positive rate of 8.29%. In addition, LR showed high precision rate of 95%, but also the highest number of false negatives of 17.04%. This error might be due to an optimized variable or numerous amount of features. As phishing e-mails always appear similar to normal e-mails, this approach might not be reliable anymore. Chandrasekaran et al. [19] proposed a technique to classify phishing based on the structural properties of phishing e-mails. They chose 25 features mixed between style markers "e.g. the security, words suspended, structural attributes, and account", such as the structure of the subject line of the e-mail and the structure of the greeting in the body. They tested 200 e-mails (100 phishing and 100 legitimate). They applied simulated annealing as an algorithm for feature selection. Then, a feature set was chosen from the dataset, and they used information gain (IG) to rank these characteristics based on their relevance. They applied a one-class SVM to classify phishing e-mails based on the selected features. Their results claim a detection rate of 95% of phishing e-mails with a low false-positive rate. In the study [20], authors applied a Bayesian classifier for detecting phishing e-mails, evaluated them in terms of accuracy, error, time, precision, and recall. The model resulted in an accuracy of 96.46%.

Rawal, Srishti et al. [15] proposes a system based on feature extraction for detection of phishing e-mails to study variety of characteristics and worked on building an approach that provides greater detection rate and uses the least amount of features. In order to achieve it, the authors extracted 9 different features from their self-made dataset based on links, tags, and words present in the body of the e-mails. They used 6 classifiers and obtain a max identity accuracy of 99.87% using RF and SVM that are supervised ML algorithms. Also, Hota, H. et al. [16] introduced a remove replace feature selection approach to categorize phished e-mails by using two Decision Tree (DT) algorithms namely Classification and Regression Tree (CART) and C4.5 and along with reducing feature subset. The ensemble model achieved an accuracy of 99.27%.They also compared and analyse their results with existing Info Gain (IG) and Gain Ration (GR) feature selection techniques. Mbah, Lashkari and Ghorbani et al. [21] proposed Phishing Alerting System (PHAS) to detect advertisement and pornographic phishing e-mails. They used WEKA and two classification algorithms: KNN and Decision Tree (J48). KNN performed better than other algorithms, and obtain the top precision and recall of 93.11%. C. EmilinShyni et al. [17] proposes a methodology incorporating ML, image processing, and NLP. They use a total of 61 features for training the prediction model. They achieved a classification accuracy of above 96% using a multi-classifier of SVM, RF, and logitboost. In the study [18], a model that utilizes 23 hybrid features of the e-mail header and body extracted from about 10000 e-mails divided equally between ham and spam e-mails is introduced, the model used J48 classifier to determine phishing and legitimate e-mails and concluded with an accuracy of 98.11% and false positive rate of 0.53%.

Recently, G. Sonowal et al. [22] suggested a binary search feature selection (BSFS) method for phishing e-mail detection, which assessed with greater accuracy using fewer features as well as less search time. The author's result shows that the weighted accuracy of the BSFS technique is 97.41% which is higher than sequential forward floating selection (SFFS - 95.63%) and wingsuit flying search (WFS - 95.56%). The study still needs more features and sophisticated feature selection techniques to get the desired best feature set. The author Y. Li et al. [27] analyzed and compared nine ML algorithms with MultiBoosting and AdaBoost algorithms for phishing detection in websites. They compared the algorithms based on F-measure, accuracy and area under the receiver operating characteristic (ROC) curves (AOC). SVM with AdaBoost performed better than other classification algorithms "CART, Rotation Forest, REP Tree, Random Tree, ANN, RF, C4.5,", and it achieved a classification accuracy of 97.61%, F-measure of 97.6 %, and ROC area of 99.6% respectively.

6. PHISHING E-MAIL DETECTION USING DEEP LEARNING

Among other machine learning phished e-mail classification techniques, M. Jameel et al. [23] used a larger dataset of 6000 e-mails for training and 3100 for testing. Visual Basic.Net programming language was used for the extraction of 18 binary features that are most visible in header and HTML body of the e-mails. They applied feed forward neural network for classifying phish e-mails from ham e-mails based on factors like e-mail features and 5 hidden neurons. An accuracy of 98.72% and a learning rate of 0.01 was obtained. They concluded that the training time was slightly greater than the testing time of algorithms for detection of the e-mails. A paradigm aiming criminals based on the character level convolutional neural network (CNN) was accomplished by some researchers [25]. The proposed paradigm focuses on URLs to extricate characteristics of e-mails that entirely remove the concept of handcrafted characteristics. The paradigm does not depend on network access, which evinces it extra trustworthy for victims owed to least reaction period. It has an accuracy of 95.02%, yet this paradigm has few disadvantages. The foremost downside is that it does not recognize if the URLs of the online sites are working properly or has some fault. It is truly significant to check URL of online sites before any major conclusion. The paradigm occasionally misidentifies phished web pages in case of shorter URLs or URLs containing confidential words like "login" or "registered", that lead to misidentification of URLs as phished online sites. Certain URLs of deceptive online sites which are not really duplicates of other online sites can go un-noticed by the paradigm relying upon the URL string. The paradigm runs on recurrent convolutional neural network (RCNN) including multilevel vectors as an attention mechanism, which allows concurrent modelling of an e-mail at the word levels, header, character, and body.

THEMIS is an innovative DL model for detecting phished e-mails [24]. THEMIS follows a mechanism of recurrent neural networks along with multilevel vectors and has an accuracy of 99.848% according to the outcomes of survey. The only flaw of the model is that it cannot detect phishing in e-mails with an e-mail body but no e-mail header. Recently, the study by authors in [28] shows the importance of semantic analysis in classifying phishing e-mails. The evaluation of various ML and DL models was performed using one-hot encoding for pre-processing the data. From an analysis of the hyper-parameters. For the CNN model, the best model was obtained for a filter size of 7, context window of 100, embedding window of 80, and pooling size of 4. It obtain an identification accuracy of 95.97%. "For Long-Short Term Memory (LSTM) hidden nodes were changed to determine the best accuracy by including and not-including dropouts. LSTM performed better with dropouts. But, for CNN with one hot encoding, the accuracy did not improve with dropouts, while for CNN with Word Embedding, the accuracy

generally improved slightly with dropouts. Using one-hot encoding, CNN performed the best of all paradigms".

7. SURVEY RESULT

Based on the criteria's i.e. dataset used, method proposed and accuracy of the proposed method by various researchers following table is generated that shows comparative analysis and assessment of various phishing e-mail detection techniques mentioned above.

Table 1.Table showing comparison and analysis of various phishing e-mail detection methods

Reference	Dataset Used	Method proposed	Classification Accuracy
[13]	6950 normal and 860 phished e-mails	PILFER(LIBSVM - A publically available support vector machine library)	99%
[14]	1700 phishing mails and 1700 legitimate mails	**6 Classifiers** - Logistic Regression (LR), Classification and Regression Trees (CART), Bayesian Additive Regression Trees (BART), Support Vector Machines (SVM), Random Forests (RF), and Neural Networks (NNet)	95.11%
[19]	100 phishing and 100 legitimate	one-class Support Vector Machine	95%
[20]	2500 mails for training and 2100 mails for testing	Bayesian classifier (based on Naïve Bayes algorithm)	96.46%
[15]	1605 phished and 414 ham e-mails	Random Forest and SVM classifier	99.87%
[16]	Publically available dataset	Remove Replace Feature Selection Technique (RRFST) using C4.5 and Classification and Regression Tree(CART) algorithm	99.27%
[21]	6951 legitimate and 2357 phishing e-mails.	KNN and Decision Tree (J48)	93.11%
[17]	5260 e-mails	Multi classifier - SVM, Random Forest, Logitboost	96.3%
[18]	5000 phishing and 5000 ham e-mails	J48 classification algorithm	98.11%
[22]	1824 phishing and 1604 legitimate e-mails	binary search feature selection	97.41%
[27]	Publically available dataset	SVM with AdaBoost algorithm	97.61%
[23]	Training phase, 6000 e-mails (3000 phished e-mails and 3000 ham e-mails) were used. In testing phase, 3100 e-mails (1550 phish e-mails and 1550 ham e-mails)	Feed Forward Neural Network	98.72%
[25]	Publically available dataset	Recurrent convolutional neural network	95.02%
[24]	Combination of various publically available dataset	THEMIS model based on Recurrent Convolutional Neural Networks	99.848%
[28]	3,416 phishing e-mails and 14,950 regular e-mails	Convolutional neural network	95.97%

8. CONCLUSION

The paper presents a survey analysis of actual phishing email identification works from various perspectives. This survey is unique in the sense that it relates works to their openly available tools and resources.Many ML methods have been adopted to identify phishing emails, but these cannot effectively detect new phishing scams, which needs significant manual feature engineering.Anti-phishing technology developed on the source code features is quite slow in terms of the classification of phishing emails given its dependence on third-party services and scraping of the email content. The analysis of the presented works revealed that not much work had been performed on phishing email detection using natural level Natural Language Processing (NLP) techniques. Therefore, many open issues are associated with this phishing email detection.

9. FUTURE SCOPE

The outcomes shows that further work is required to employ modernized DL techniques in phishing email detection studies, for instance, Convolutional Neural networks (CNN), Recurrent Neural Networks (RNNs), and Deep Reinforcement Learning models. In the last few years, phishing emails have been increasing at unprecedented levels and the counter measures used against this evolving threat have not proven effective despite their constant upgrade and revision. To prevent this threat of phishing emails, more advanced phishing detection technology is necessary. The tools and resources are not sufficient in this research area. Hence, the researchers are in dire need to perform more research efforts to assess DL techniques in the phishing email detection domain.

10. REFERENCES

[1] Ş. Şentürk, E. Yerli and İ. Soğukpınar, "E-mail phishing detection and prevention by using data mining techniques", 2017 International Conference on Computer Science and Engineering (UBMK), pp. 707-712.

[2] Anti-Phishing Working Group. 2021. APWG Phishing Activity Trends Report [Online]. Available: https://docs.apwg.org/reports/apwg_trends_report_q1_2021.pdf

[3] Z Alkhalil, C Hewage, L Nawaf, I Khan, "Phishing Attacks: Recent Comprehensive Study and a New Anatomy - Frontiers in Computer Science", 2021.

[4] Hong J., "The state of phishing attacks", 2012, Communications of the ACM, vol. 55(1), pp. 74-81.

[5] Ibrahim, N. Al Herami, E. Al Naqbi, M. Aldwairi, "Detection and Analysis of Drive-by-Downloads and Malicious Websites", 2019, Seventh International Symposium on Security in Computing and Communications (SSCC'19), Trivandrum, Kerala, India.

[6] Masri, R., Aldwairi, M. Automated malicious advertisement detection using VirusTotal, URLVoid, and TrendMicro. 8th International Conference on Information and Communication Systems (ICICS), Irbid, 336- 341, (2017). doi:10.1109/IACS.2017.7921994. https://doi.org/10.1109/ IACS.2017.7921994

[7] A. Hamid I.R., Abawajy J., "Hybrid Feature Selection for Phishing E-mail Detection", 2011, International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP 2011), vol. 7017. Springer, Berlin, Heidelberg.

[8] Aburrous, M., Hossain, M. A., Thabatah, F. and Dahal, K. P., 2008, "Intelligent phishing website detection system using fuzzy techniques", 3rd International Conference on Information & Communication Technologies: From Theory to Applications (ICCTA'08). New York: IEEE

[9] G. Varshney, M. Misra, and P. K. Atrey, 2016, "A survey and classification of web phishing detection schemes," Secur. Commun. Networks, vol. 9, no. 18, pp. 6266–6284.

[10] Vijayalakshmi, M., Mercy Shalinie, S., Yang, M. H., & U., R. M., 2020, "Web phishing detection techniques: a survey on the state-of-the-art, taxonomy and future directions", IET Networks, vol. 9(5), pp. 235–246.

[11] Said Salloum, TarekGaber, Sunil Vadera, Khaled Shaalan, 2021, "Phishing E-mail Detection Using Natural Language Processing Techniques: A Literature Survey", Procedia Computer Science, vol. 189,2021, pp. 19-28.

[12] Alsariera, Y.A., Adeyemo, V.E., Balogun, A.O., Alazzawi, A.K., 2020, "AI Meta-Learners and Extra-Trees Algorithm for the Detection of Phishing Websites". IEEE Access 2020, vol. 8, pp. 142532–142542.

[13] Fette, I., Sadeh, N., Tomasic, A., 2006, "Learning to Detect Phishing E-mails", Technical report, Institute of Software Research International, School of Computer Science, Carneige Melon University.

[14] Abu-Nimeh, S., Nappa, D., Wang, X., Nair, S., 2007, "Comparison of Machine Learning Techniques for Phishing Detection". APWG eCrime Researchers Summit, Pittsburgh, USA.

[15] Rawal, Srishti&Rawal, Bhuvan&Shaheen, Aakhila& Malik, Shubham, 2017. "Phishing Detection in E-mails using Machine Learning. International Journal of Applied Information Systems". 12, pp. 21-24.

[16] Hota, H. &Shrivas, A.K. &Hota, Rahul, 2018, "An Ensemble Model for Detecting Phishing Attack with Proposed Remove-Replace Feature Selection Technique". Computer Science. Vol. 132.pp. 900-907.

[17] C. EmilinShyni, S. Sarju, S. Swamynathan, 2016, "A Multi-Classifier Based Prediction Model for Phishing E-mails Detection Using Topic Modelling, Named Entity Recognition and Image Processing". Circuits and Systems, vol. 07, pp. 2507-2520.

[18] Sami Smadi, NaumanAslam, Li Zhang, RafeAlasem, M A Hossain, "Detection of Phishing E-mails using Data Mining Algorithms", 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), 2015.

[19] Chandrasekaran M., Narayanan K. and Upadhyaya S., 2006, "Phishing E-mail Detection Based on Structural Properties", first annual Symposium on Information Assurance: Intrusion Detection and Prevention, New York, pp. 2-8

[20] Sunil B. Rathod, Tareek M. Pattewar, 2015, "Content Based Spam Detection in E-mail using Bayesian Classifier", IEEE ICCSP conference.

[21] Mbah, K. F. Lashkari, A. H., Ghorbani, A. A. "A phishing e-mail detection approach using machine learning techniques", World Academy of Science, Engineering and Technology, Computer and Information Engineering, vol. 3(1), pp. 2333,

[22] G. Sonowal, "Phishing E-mail Detection Based on Binary Search Feature Selection", 2020, SN Computer. Science, vol. 1.

[23] M. Jameel, Noor & George, Loay.,,"Detection of Phishing E-mails using Feed Forward Neural Network"., 2013 International Journal of Computer Applications.

[24] Y. Fang, C. Zhang, C. Huang, L. Liu and Y. Yang, 2019, "Phishing E-mail Detection Using Improved RCNN Model With Multilevel Vectors and Attention Mechanism," in IEEE Access, vol. 7, pp. 56329-56340.

[25] A. Aljofey, Q. Jiang, Q. Qu, M. Huang, and J.-P. Niyigena, 2020, "An Effective Phishing Detection Model Based on Character Level Convolutional Neural Network from URL," Electronics, vol. 9, no. 9, p. 1514.

[26] Suman Mann etal. "Smart hospital using AI and IOT for covid-19 Pandemic" Smart Healthcare Monitoring Using IoT with 5G Challenges, Directions, and Future Predictions (1st ed.). CRC Press. https://doi.org/10.1201/9781003171829

[27] Y. Li, Z. Yang, X. Chen, H. Yuan, and W. Liu, 2019,"A stacking model using URL and HTML features for phishing webpage detection," Future Gener. Comput Syst., vol. 94, pp. 27–39.

[28] Bagui, S., Nandi, D. & White, R. J. , 2021, "Machine Learning and Deep Learning for Phishing E-mail Classification using One-Hot Encoding. Journal of Computer Science", vol. 17(7), pp. 610-623.

[29] Bagui, S., Nandi, D. & White, R. J. , 2021, "Machine Learning and Deep Learning for Phishing E-mail Classification using One-Hot Encoding. Journal of Computer Science", vol. 17(7), pp. 610-623.

[30] P. Manojkumar, M. Suresh, Alim Al Ayub Ahmed, Hitesh Panchal, Christopher AsirRajan, A. Dheepanchakkravarthy, A. Geetha, B. Gunapriya, Suman Mann &Kishor Kumar Sadasivuni (2021) A novel home automation distributed server management system using Internet of Things, International Journal of Ambient Energy, DOI: 10.1080/01430750.2021.1953590.

[31] Akshay Chopra, Bhavya Chaudhary, Suman Mann, "Analysis of Security Issues in VoIP" International journal of computer application", Volume 103 – No.8, October 2014

[32] Suman Mann, Tanya Jain, Aakash Vyas, "The Blockchain Revolution: Paradigm Shifts in Traditional Voting Practices", International journal of computer application, Volume 176 – No. 37, July 2020.

[33] Anish Batra, Guneet Singh Sethi, Suman Mann," Personalized Automation of Electrical and Electronic Devices Using Sensors and Artificial Intelligence—the Intelligizer System" Computational Intelligence: Theories, Applications and Future Directions - Volume I, 2019, Volume 798

[34] SakshiHooda, Suman Mann, "Sepsis-Diagnosed Patients' In-Hospital Mortality Prediction Using Machine Learning: The Use Of Local Big Data-Driven Technique In The Emergency Department" International journal of grid and distributed computing, vol13, Issue 1 2020

[35] Kaushik, Anupama&Tayal, Devendra&Yadav, Kalpana. "The Role of Neural Networks and Metaheuristics in Agile Software Development Effort Estimation", 2020, International Journal of Information Technology Project Management. 11. 50-71. 10.4018/IJITPM.2020040104.

[36] Arora J., Tushir M. (2021) Performance Analysis of Different Kernel Functions for MRI Image Segmentation. In: Bansal P., Tushir M., Balas V., Srivastava R.(eds) Proceedings of International Conference on Artificial Intelligence and Applications. Advances in Intelligent Systems and Computing, vol 1164. Springer, Singapore.

[37] Jyoti Khurana, Vachali Aggarwal and Harjinder Singh.,"A Comparative Study of Deep Learning Models for Network Intrusion Detection", International Journal of Computer Applications, 174(23):38-46, March 2021.