International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Enhancement of Text Recognizing Exploitation in Phishing Websites using LSTM in Comparison with CNN based on Improving the Accuracy Rate

Shaik Yakub Pasha

Student, Kl University

Abstract

Aim: The objective of the work is to predict the accuracy of phishing websites based on exploitation of text recognition using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM). To achieve accuracy, a novel np.random function was used.

Method and Materials: Accuracy and Loss are performed with DATA dataset from the keras library. The total sample size is 20. The two groups Convolutional Neural Network (N=10) and Long Short Memory (N=10).

Result: The result proved that Convolutional Neural Network (CNN) with better accuracy of 97.3% than Long Short Term Memory (LSTM) accuracy of 93.2%. Finally CNN appears significantly better than LSTM. The two algorithms CNN and LSTM are statistically satisfied with the independent sample T-Test value ($p<0.001$) with a confidence levelof 95%.

Conclusion: Detecting the phishing website significantly seems to be better in CNN (Std.Error Mean = .0632) than LSTM (Std.Error Mean = .0678).

Keywords: Phishing Websites Detection, Deep Learning, Convolutional Neural Network, Long Short Term Memory, Novel np.random function.

Introduction

Phishing is a cyber attack that uses disguised email or message as a weapon. Since most of the users go online to access the services provided by the government and financial companies, there has been a significant increase in phishing attacks for the past years. Various methods are used by phishers to attack the users such as messaging, spoofed links and fake websites. The reason for creating these websites is to get the data from users like account numbers, login details, passwords of debit and credit card details, etc. (Mao et al. 2018).Importance in today's world security will be increased and reduce the risk of information. It can be used in various areas like banking, education sites E-commerce and paytm, etc (Ezekiel, Taylor, and Deedam- Okuchaba 2020); (Mohammad, McCluskey, and Thabtah 2013). The intelligence report states that phishing grew by 40% in 2019 and noticed assaults focusing on accreditations for money, email, cloud, etc.

Phishing website detection is carried out by researchers 20 related research articles in IEEE

digital Xplore and 35 articles are published in the research gate. Analyzed the detection phishing website using the Support vector machine (SVM) and some Anomaly features (Pan and Xuhua 2006). Used 'CANTINA' content based technique to detect phishing websites using the term-frequency-inverse-frequency (Zhang, Hong, and Cranor 2007). To reduce the feature selection model to detect phishing websites using Logistic Regression (LR) and Support Vector Machine (SVM) (Fadheel, Abusharkh, and Abdel-Qader 2017b). Discussed making a comparative study to detect malicious URL with classical machine learning techniques – logistic regression using bigram, deep learning techniques like convolution neural network (CNN) and CNN long short-term memory (CNN-LSTM) as architecture. (Vazhayil, Ravi, and Kp 2018). Used K-Nearest neighbor (KNN), Support Vector Machine (SVM) and Random Forest to detect phishing websites (Desai et al. 2017).

Previously our team had rich experience in working on various research projects across multiple disciplines. (S. Vigneshwaran et al. 2020; Kumaran et al. 2020; Manickam et al. 2019; Shanmugam et al. 2021; Kaliaperumal Rukmani et al. 2020; Lakshmi and Hemamalini 2019; Anitha Kumari and Srinivasan 2019; Saravanan et al. 2021; Rajkumar and Ganapathy 2020; Shanmugam Vigneshwaran et al. 2021; Sivakumar, Anbalagan, and Jayavel 2020; Stephen Leon J, Bharathiraja G, and Jayakumar 2020; Balasaraswathi et al. 2020; Anitha, Naresh, and Devi 2020; Sivasamy et al. 2021). Now the growing trend in this area motivated us to pursue this project.

Based on literature surveys, LSTM has the least accuracy. Detection of phishing is shownto be a very low percentage while analyzing the websites and the manual input is not possible to add to the dataset. The aim of the study is to improve the accuracy of the phishing website, andto reduce the loss of the data while training and testing the dataset. The novel np.random function used to achieve accuracy.

Materials and Methods

The study setting of the proposed work is done in Saveetha School of Engineering.. Two groups were identified for the study setting where group one CNN and group two LSTM. Using G power 10 sample sizes and totally 20 sample sizes have been carried out for our study, 95% confidence and pretest power 80% (Adeyemo et al. 2021).

The dataset named 'DATA' is downloaded from the public domain keras library. In our experiments here we used the data.csv dataset. Detailed description of the features/attributes in the dataset can be found below in the form of a Table 1. The dataset consists of 5 lakhs instances and contains url and label. The dataset was splitted into two parts namely the training part and testing part. 70% of the data was used for training and the remaining 30% was used for testing.

The algorithm was implemented by evaluating the train and test. Input dataset collected from thislink (antonyj n.d.).

Framework of Text Recognition in Phishing Website

The input is given, records that are present in the dataset and reading the dataset into the program. Using Convolutional Neural Network and Long Short Term Memory, the input is processed

and the data is divided into training and testing parts. LSTM, Sequential() is used to train the data and the np.random of CNN is used to add hidden layers that are connected to the input layer and train the data. The function y_pred which is present in LSTM and np.random is used to predict the output of the testing data. The predicted output is the phishing website and it is displayed as either good or bad and accurate. The framework of predicting the text recognition in phishing websites step by step is shown in Fig. 1.

Convolutional Neural Network (CNN) Algorithm

CNN is a deep learning technique that works well for identifying simple patterns in the data which will then be used to form more complex patterns in subsequent layers (Yerima and Alzaylaee 2020). Two types of layers are used for building CNNs; convolutional layers and pooling layers. The part of the convolutional layer is to recognize neighborhood conjunctions of highlights from the past layer, while the job of the pooling layer is to combine semantically comparable highlights into one. Pseudo code for CNN algorithm is shown in Table 2. The accuracy of Convolutional Neural Network (CNN) with changing the test size shown in Table 6.

Long Short Term Memory (LSTM) Algorithm

Long Short Term Memory (LSTM) networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems(Pooja, Lakshmi, and Sridhar 2020). Mostly used in machine translation, speech recognition, and more. It can be hard to get your hands around what LSTMs are, and how terms like bidirectional and sequence-to-sequence relate to the field. An LSTM layer consists of a set of recurrently connected blocks, known as memory blocks. Pseudo code for LSTM algorithm is shown in Table
3. The accuracy of Long Short Term Memory (LSTM) with changing the test size shown in Table 7. Equation (1) shows the accuracy measurement.

Accuracy=((TP+TN)/(TP+TN+FP+FN))---------- (1)

Where,
TP - is the no. of true positives classified by the model.FP - is the no.of false positive classified by the model.

TN - is the no.of true negative classified by the model. FN - is the no.of false negative classified by the model.

The software tool to evaluate CNN and LSTM algorithms is colab in Python programming language. The hardware configuration includes an intel i3 processor with a RAM size of 4GB. The system used was a 64-bit windows 10 operating system.

Statistical Analysis

For statistical implementation, the software to be used here is IBM SPSS V26.0. Statistical package for social sciences is used for calculating the statistical calculations such as mean, standard deviation, and also to plot the graphs etc.,. The independent variables are Url, Label and the dependent

variable is 'accuracy'. In SPSS, the dataset is prepared using 10 as sample size for each group and accuracy is given as the testing variable.

Results

For convolutional neural network (CNN) and Long Short Term Memory (LSTM) and compared the both algorithms with their accuracy rate. For both proposed and existing algorithms 10 iterations were taken for each iteration the predicted accuracy was noted for analyzing accuracy. The results of statistical packages of social sciences (IBM-SPSS v21) used for data analysis. With value obtained from the iterations Independent Sample T-test was performed . Significance values and group statistics values of proposed and existing algorithms are shown in Table 4 & Table 5. Whereas t-test equality is calculated. Confidence interval of the difference as lower and upper values range as shown in Table 5.

The bar graph is plotted by selected mean accuracy on Y-axis and the Group on X-axis. From the graph, it is clear that CNN has significantly higher accuracy than LSTM shows in Fig. The error bars are shown in the graph and the error rate is less for CNN compared to LSTM.

Discussions

In this study, the CNN algorithm has better significant phishing website prediction accuracy tha LSTM algorithm ($p<0.001$, Independent sample t-test). The improved accuracy and reduced loss for CNN (Accuracy = 97.40%, Loss = 3.6%) than LSTM (accuracy = 93.16%, Loss= 6.8%).

Various machine learning algorithms are used to predict phishing websites. Proposed a reduced feature selection model to detect phishing websites. They used Logistic Regression and Support Vector Machine (SVM) as classification methods to validate the feature selection method. 30 site features have been selected and used for phishing detection. The LR and SVM calculations performance was surveyed dependent on precision, recall, f-measure and accuracy. Study shows that SVM algorithm achieved best performance over LR algorithm (Fadheel,Abusharkh, and Abdel-Qader 2017a). Proposes a classification model in order to classify the phishing attacks. This model contains feature extraction and classification of websites. In feature extraction, features have been taken from a dataset. To classify these features, SVM, NB and ELM were used. In this activation functions were used and achieved 95.34% accuracy than SVM and NB. With the help of MATLAB results are obtained (Sonmez et al. 2018). Proposed a flexible filtering decision module to extract features automatically without any specific expert knowledge of the URL domain using neural network model. In this methodology the author utilized every one of the characters remembered for the url strings and byte esteems. They not just check byte esteems and furthermore cover portions of adjoining characters by moving 4-bits. They embed combination information of two characters appearing sequentially and count how many times each value appears in the original URL string and achieve a 512 dimension vector. These are factors affecting the phishing website gender, age and exposure to educational materials all have some impact etc (Shima et al. 2018). Compared five machine learning algorithms and calculated the accuracy, precision, recall and f1 score. By this value the author decided on a significantly better algorithm. Random Forest given the higher accuracy 87%. (Parasar and Jadhav 2021). Classification model was proposed to solve the difficulties encountered in phishing websites detection using the logistic regression, SVM and Naive Bayes.SVM has higher accuracy 87% obtained (Satheeshkumar 2019).

Our institution is passionate about high-quality evidence-based research and has excelled in various fields (Vijayashree Priyadharsini 2019; Ezhilarasan, Apoorva, and Ashok Vardhan 2019; Ramesh et al. 2018; Mathew et al. 2020; Sridharan et al. 2019; Pc, Marimuthu, and Devadoss 2018; Ramadurai et al. 2019). We hope this study adds to this rich legacy.

The limitation in our study is that as phishing websites increase day by day, some featuresmay be included or replaced with new ones to detect them. In the future, to improve the model training process by automatically the search and selection of the key parameters (i.e. number of filters, filter lengths, and number of fully connected units) that jointly results in the optimal performing CNN model.

Conclusion

In this research work, the results indicate that our proposed Convolutional Neural Network (CNN) based model with novel np.random function can be used to detect previously unseen phishing websites with improved accuracy of 97%.

Declarations

Conflict of Interests

The authors do not have any conflict of interest associated with this manuscript.

Acknowledgement

The authors would like to express their gratitude towards Tetra Institute for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organization for providing financial support that enabled us to complete the study.

1. Tetra Mind
2. KL University

References

1. Adeyemo, Victor, Abdullateef Balogun, Hammed Mojeed, and K. S. Adewole. 2021. "Ensemble-Based Logistic Model Trees for Website Phishing Detection." *Advances in Cyber Security*, February, 627–41.
2. Anitha, K., K. Naresh, and D. Rukmani Devi. 2020. "A Framework to Reduce Category Proliferation in Fuzzy ARTMAP Classifiers Adopted for Image Retrieval Using Differential Evolution Algorithm." *Multimedia Tools and Applications* 79 (5-6): 4217–38.
3. Anitha Kumari, S., and S. Srinivasan. 2019. "Ash Fouling Monitoring and Soot-Blow Optimization for Reheater in Thermal Power Plant." *Applied Thermal Engineering* 149 (February): 62–72.
4. antonyj. n.d. "Malicious_n_Non-Malicious URL." Accessed March 30, 2021. https://www.kaggle.com/antonyj453/urldataset.
5. Balasaraswathi, M., Mehtab Singh, Jyoteesh Malhotra, and Vigneswaran Dhasarathan. 2020. "A High-Speed Radio-over-Free-Space Optics Link Using Wavelength Division Multiplexing-Mode Division Multiplexing-Multibeam Technique." *Computers & Electrical Engineering* 87 (106779):

106779.
6. Desai, Anand, J. Jatakia, Rohit Naik, and Nataasha Raul. 2017. "Malicious Web Content Detection Using Machine Leaning." http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8256834.
7. Ezekiel, P. S., O. E. Taylor, and F. B. Deedam- Okuchaba. 2020. "A Model to Detect Phishing Websites Using Support Vector Classifier and a Deep Neural Network Algorithm." *IJARCCE* 9 (6): 188–94.
8. Ezhilarasan, Devaraj, Velluru S. Apoorva, and Nandhigam Ashok Vardhan. 2019. "Syzygium Cumini Extract Induced Reactive Oxygen Species-Mediated Apoptosis in Human Oral Squamous Carcinoma Cells." *Journal of Oral Pathology & Medicine: Official Publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology* 48 (2): 115–21.
9. Fadheel, Wesam, Mohamed Abusharkh, and Ikhlas Abdel-Qader. 2017a. "On Feature Selection for the Prediction of Phishing Websites." *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*.
10. https://doi.org/10.1109/dasc-picom-datacom-cyberscitec.2017.146.
11. ———. 2017b. "On Feature Selection for the Prediction of Phishing Websites." *2017 IEEE 15thIntl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*,November, 871–76.
12. Kaliaperumal Rukmani, Devabalaji, Yuvaraj Thangaraj, Umashankar Subramaniam, Sitharthan Ramachandran, Rajvikram Madurai Elavarasan, Narottam Das, Luis Baringo, andMohamed Imran Abdul Rasheed. 2020. "A New Approach to Optimal Location and Sizing of DSTATCOM in Radial Distribution Networks Using Bio-Inspired Cuckoo Search Algorithm." *Energies* 13 (18): 4615.
13. Kumaran, P., S. Mohanamurugan, S. Madhu, R. Vijay, D. Lenin Singaravelu, A. Vinod, M. R. Sanjay, and Suchart Siengchin. 2020. "Investigation on Thermo-Mechanical Characteristics of Treated/untreated Portunus Sanguinolentus Shell Powder-Based Jute Fabrics Reinforced Epoxy Composites." *Journal of Industrial Textiles* 50 (4): 427–59.
14. Lakshmi, M., and S. Hemamalini. 2019. "Coordinated Control of MPPT and Voltage Regulation Using Single-Stage High Gain DC–DC Converter in a Grid-Connected PV System." *Electric Power Systems Research* 169 (April): 65–73.
15. Manickam, Adhiyaman, Ezhilmaran Devarasan, Gunasekaran Manogaran, Malarvizhi Kumar Priyan, R. Varatharajan, Ching-Hsien Hsu, and Raja Krishnamoorthi. 2019. "Score Level Based Latent Fingerprint Enhancement and Matching Using SIFT Feature." *Multimedia Tools and Applications* 78 (3): 3065–85.
16. Mao, Jian, Jingdong Bian, Wenqian Tian, Shishi Zhu, and Zhenkai Liang. 2018. "Detecting Phishing Websites via Aggregation Analysis of Page Layouts." *Procedia Computer Science* 129 (January): 224–30.
17. Mathew, M. G., S. R. Samuel, A. J. Soni, and K. B. Roopa. 2020. "Evaluation of Adhesion of Streptococcus Mutans, Plaque Accumulation on Zirconia and Stainless Steel Crowns, and Surrounding Gingival Inflammation in Primary …." *Clinical Oral Investigations*. https://link.springer.com/article/10.1007/s00784-020-03204-9.

18. Mohammad, Rami, T. L. McCluskey, and Fadi Abdeljaber Thabtah. 2013. "Predicting Phishing Websites Using Neural Network Trained with Back-Propagation," January. http://dx.doi.org/.
19. Pan, Ying, and Ding Xuhua. 2006. "Anomaly Based Web Phishing Page Detection." *22nd Annual Computer Security Applications Conference (ACSAC 2006), 11-15 December 2006, Miami Beach, Florida, USA*, December, 381–92.
20. Parasar, Deepa, and Yogesh H. Jadhav. 2021. "An Automated System to Detect Phishing URL by Using Machine Learning Algorithm." *International Conference on Mobile Computing and Sustainable Informatics*, January, 217–25.
21. Pc, J., T. Marimuthu, and P. Devadoss. 2018. "Prevalence and Measurement of Anterior Loop ofthe Mandibular Canal Using CBCT: A Cross Sectional Study." *Clinical Implant Dentistry and Related Research*. https://europepmc.org/article/med/29624863.
22. Pooja, A. S. S. V. Lakshmi, A. S. S. Lakshmi, and M. Sridhar. 2020. "Analysis of Phishing Website Detection Using CNN and Bidirectional LSTM." *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. https://doi.org/10.1109/iceca49313.2020.9297395.
23. Rajkumar, R., and Velappa Ganapathy. 2020. "Bio-Inspiring Learning Style Chatbot Inventory Using Brain Computing Interface to Increase the Efficiency of E-Learning." *IEEE Access: Practical Innovations, Open Solutions* 8: 67377–95.
24. Ramadurai, Neeraja, Deepa Gurunathan, A. Victor Samuel, Emg Subramanian, and Steven J. L. Rodrigues. 2019. "Effectiveness of 2% Articaine as an Anesthetic Agent in Children: Randomized Controlled Trial." *Clinical Oral Investigations* 23 (9): 3543–50.
25. Ramesh, Asha, Sheeja Varghese, Nadathur D. Jayakumar, and Sankari Malaiappan. 2018. "Comparative Estimation of Sulfiredoxin Levels between Chronic Periodontitis and Healthy Patients - A Case-Control Study." *Journal of Periodontology* 89 (10): 1241–48.
26. Saravanan, A., P. Senthil kumar, Dai-Viet N. Vo, S. Jeevanantham, V. Bhuvaneswari, V. Anantha Narayanan, P. R. Yaashikaa, S. Swetha, and B. Reshma. 2021. "A Comprehensive Review on Different Approaches for CO2 Utilization and Conversion Pathways." *Chemical Engineering Science* 236 (116515): 116515.
27. Satheeshkumar, Akila D. 2019. "Phishing Websites Detection Using Machine Learning" 8 (2S11): 111–14.
28. Shanmugam, Vigneshwaran, Oisik Das, Karthik Babu, Uthayakumar Marimuthu, Arumugaprabu Veerasimman, Deepak Joel Johnson, Rasoul Esmaeely Neisiany, Mikael S. Hedenqvist, Seeram Ramakrishna, and Filippo Berto. 2021. "Fatigue Behaviour of FDM-3D Printed Polymers, Polymeric Composites and Architected Cellular Materials." *International Journal of Fatigue* 143 (106007): 106007.
29. Shima, Keiichi, Daisuke Miyamoto, Hiroshi Abe, Tomohiro Ishihara, Kazuya Okada, Yuji Sekiya, Hirochika Asai, and Yusuke Dois. 2018. "Classification of URL Bitstreams Using Bag of Bytes." *2018 21st Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN)*. https://doi.org/10.1109/icin.2018.8401597.
30. Sivakumar, N., G. Anbalagan, and R. Jayavel. 2020. "Crystal Design, Thermal and Dielectric Behavior of Novel Silver (Ag) Co-Ordinated Thiourea Single Crystals." *Materials Letters* 272 (127899): 127899.
31. Sivasamy, Ramesh, Marutheeswaran Srinivasan, Rodrigo Espinoza-González, and Edgar Mosquera.

2021. "Electronic and Optical Studies on Two-Dimensional Hydrogenated Stirrup Triels Nitride Nanosheets: A First-Principle Investigation." *Materials Science & Engineering. B, Solid-State Materials for Advanced Technology* 264 (114978): 114978.

32. Sonmez, Yasin, T. Tuncer, Hüseyin Gökal, and Engin Avci. 2018. "Phishing Web Sites Features Classification Based on Extreme Learning Machine." *2018 6th International Symposium on Digital Forensic and Security (ISDFS)*, March, 1–5.
33. Sridharan, Gokul, Pratibha Ramani, Sangeeta Patankar, and Rajagopalan Vijayaraghavan. 2019. "Evaluation of Salivary Metabolomics in Oral Leukoplakia and Oral Squamous Cell Carcinoma." *Journal of Oral Pathology & Medicine: Official Publication of theInternational Association of Oral Pathologists and the American Academy of Oral Pathology* 48 (4): 299–306.
34. Stephen Leon J, Bharathiraja G, and Jayakumar. 2020. "Analytical and Experimental Investigations of Optimum Thermomechanical Conditions to Use Tools with Non-Circular Pin in Friction Stir Welding." *International Journal of Advanced Manufacturing Technology* 107 (11-12): 4925–37.
35. Vazhayil, Anu, Vinayakumar Ravi, and Soman Kp. 2018. "Comparative Study of the Detection of Malicious URLs Using Shallow and Deep Networks." *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, July, 1–6.
36. Vigneshwaran, Shanmugam, K. M. John, R. Deepak Joel Johnson, Marimuthu Uthayakumar, V. Arumugaprabu, and Sundaresan T. Kumaran. 2021. "Conventional and Unconventional Machining Performance of Natural Fibre-Reinforced Polymer Composites: A Review." *Journal of Reinforced Plastics and Composites* 40 (15-16): 553–67.
37. Vigneshwaran, S., R. Sundarakannan, K. M. John, R. Deepak Joel Johnson, K. Arun Prasath, S. Ajith, V. Arumugaprabu, and M. Uthayakumar. 2020. "Recent Advancement in the Natural Fiber Polymer Composites: A Comprehensive Review." *Journal of Cleaner Production* 277 (124109): 124109.
38. Vijayashree Priyadharsini, Jayaseelan. 2019. "In Silico Validation of the Non-Antibiotic Drugs Acetaminophen and Ibuprofen as Antibacterial Agents against Red Complex Pathogens." *Journal of Periodontology* 90 (12): 1441–48.
39. Yerima, Suleiman Y., and Mohammed K. Alzaylaee. 2020. "High Accuracy Phishing Detection Based on Convolutional Neural Networks." *2020 3rd International Conference on Computer Applications & Information Security (ICCAIS).*
40. https://doi.org/10.1109/iccais48893.2020.9096869.
41. Zhang, Yue, Jason Hong, and Lorrie Cranor. 2007. "CANTINA: A Content-Based Approach to Detecting Phishing Web Sites." *Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007*, January, 639–48.

TABLES AND FIGURES

Table 1. Sample Dataset Contains URL and Label

S.no	URL	Label
1	diaryofagameaddict.com	bad
2	espdesign.com.au	bad
3	iamagameaddict.com	bad

4	kalatzis.net	bad
5	slightlyoffcenter.net	bad
6	toddscarwash.net	good
7	tubemoviez.com	bad
8	ipl.hk	bad
9	crackspider.us/toolbar/install.php?pack=exe	bad
10	pos-kupang.com/	good

Table 2. Pseudocode for Convolutional Neural Network (CNN) algorithm

// I : Input dataset records
1. Import the required packages.
2. Convert the string values in the dataset to numerical values.
3. Assign the data to X_train, y_train, X_test and y_test variables.
4. Using train_test_split() function, pass the training and testing variables and give test_size and random_state as the parameters.
5. Import the np.random()
6. Predict the output using y_pred and y_test function
7. Calculate accuracy of the model.
Output **// Accuracy**

Table 3. Pseudocode for Long Short Term Memory (LSTM) Algorithm

// I : Input dataset records
1. Import the required packages.
2. Convert the string values in the dataset to numerical values.
3. Assign the data to X_train, y_train, X_test and y_test variables.
4. Using train_test_split() function, pass the training and testing variables and give test_size andthe random_state as parameters.
5. Import the Sequential() from sklearn library.
6. Using Sequential, predict the output of the testing data.
7. Calculate the accuracy.
OUTPUT **//Accuracy**

Table 4. Group statistics results (Mean of CNN 97.40 is more Compared with LSTM 93.16 and Std.Error Mean for CNN is .0632 and CNN is .0678)

Algorithm	**N**	**Mean**	**Std. Deviation**	**Std.Error Mean**
CNN	5	97.400	.1517	.0632

LSTM	5	93.160	.1517	.0678
CNN LSTM	5. 5	3.60 7.28	.141 .327	.063 .146

Table 5. Independent Sample T- test Result is applied for dataset fixing confidence interval as 95% and level of significance as 0.05 (CNN appears to perform significantly better than LSTMwith the value of $p<0.001$)

	Levene's test for equality of variances		**t-test for Equality of Means**						
								95% confidence interval of the difference	
	F	**Sig.**	**t**	**df**	**Sig. (2-tailed)**	**Mean Difference**	**Std.Error Difference**	**Lower**	**Upper**
Accuracy Equal variances Assumed Equal variance not Assumed	.071	.797	-45.721 -45.721	8 7.961	<.001 < .001	-.4.2400 -.4.2400	.0927 .0927	-4.4539 -4.4540	-4.0251 -4.0250
Loss Equal variance Assumed Equal variance not Assumed	1.169	.311	23.090 23.090	8 5.445	<.001 < .001	3.680 3.680	.159 .159	3.312 3.280	4.048 4.048

Table 6. Accuracy of Phishing Website Detection using Convolutional Neural Network (CNN) Algorithm (Accuracy = 97.40)

Test size	Accuracy
Test 1	97.40
Test 2	96.31
Test 3	95.58
Test 4	89.22
Test 5	86.42
Test 6	85.23
Test 7	75.00
Test 8	85.00
Test 9	75.00
Test 10	82.00

Table 7. Accuracy of Phishing Website Detection using Long Short Term Memory (LSTM) Algorithm (Accuracy = 93.16)

Test Size	Accuracy
Test 1	93.16
Test 2	90.33
Test 3	85.23
Test 4	89.00
Test 5	85.02
Test 6	83.00
Test 7	81.00
Test 8	83.00
Test 9	81.00
Test 10	83.22

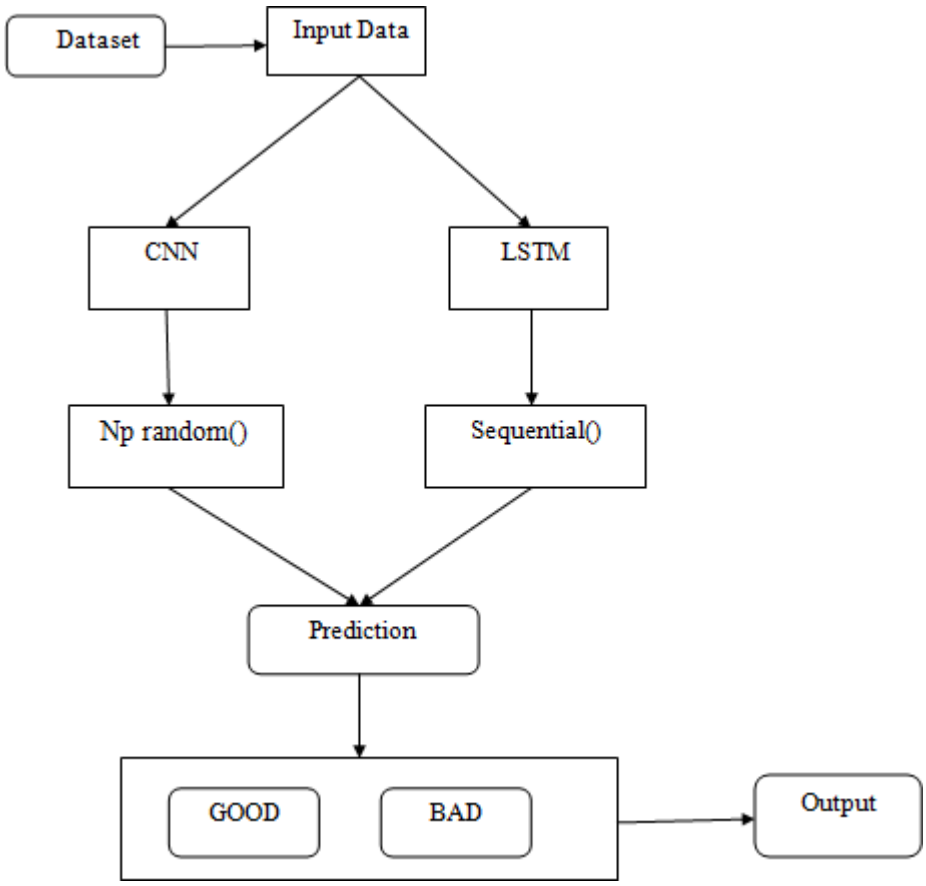

Fig. 1. Phishing Website prediction Framework using Convolutional Neural Network (CNN) andLong Short Term Memory (LSTM).

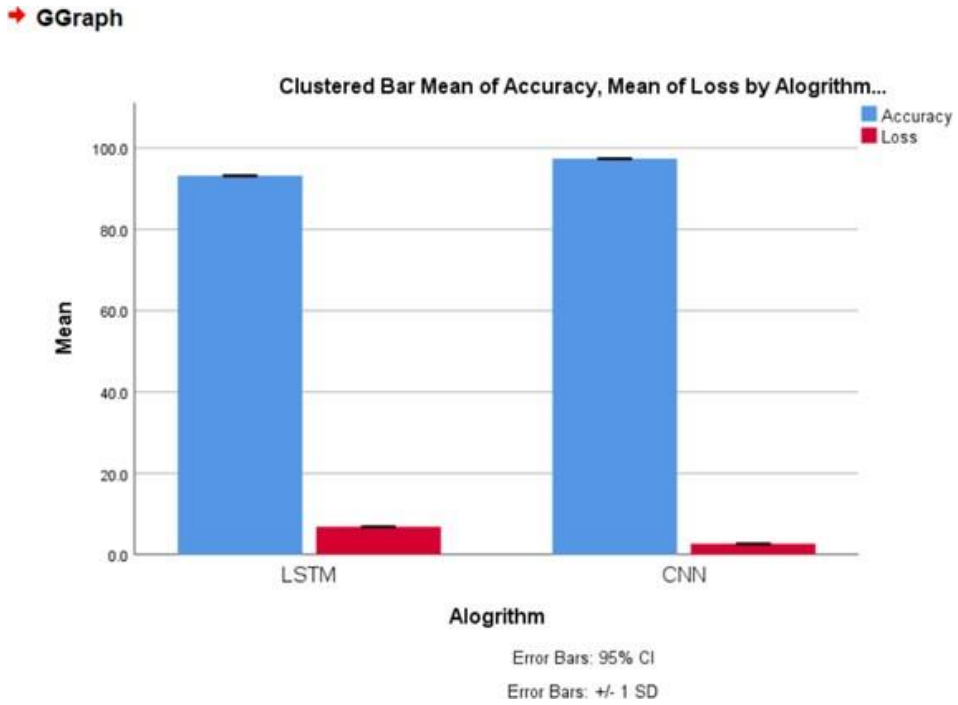

Fig. 2. Clustered Bar mean of accuracy , mean of loss by CNN & LSTM classifier in terms of mean accuracy. The mean accuracy of CNN is better than LSTM and standard deviation of CNNis slightly better than LSTM. X-axis : CNN vs LSTM algorithm Y-axis : Mean accuracy of detection ± 1 SD.