JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

PHISHING WEBSITE DETECTION USING DEEP LEARNING FRAMEWORK

Prof. A.B.Gadewar[1] ,

Vaishnavi Narkhede[2], Rutuja Mashare[3] , Amita Chorghe[4] ,Sourav Sonawane[5]

[1] Assistant Professor, Dept. of Information Technology, Pune District Education Association's College of Engineering, Pune, Maharashtra, India.

[2,3,4,5] Students of Bachelor of Engineering, Dept. of Information Technology, Pune District Education Association's College of Engineering, Pune, Maharashtra, India

Abstract : The Phishing is a cyber crime where a person who poses as a legitimate agency contacts a victim or target via email, phone or text message to attract the person to supply in-formation, information about personal identity, banking and credit card information and passwords. phishing is a crime. The new term 'fishing' refers to the attacker's invitation to visit a counterfeit site by creating a website look, and to get personal information from users such as username, password, financial information, account details, national security identifier, etc.. Phishing is a new term that was developed using 'fishing.

The information collected is used for potential target ads or even identity robberies, attacks (for example, money transfer from one's account). The attack method that is widely used is to send e-mails, messages that can lead to data theft or personal information. Social networking account Passwords, credit cards or attackers provide upgrades to their websites, encourage you to comply with your personal information and change it via fake website, are mis-entered daily. If you are entering your personal data, the attackers will collect it successfully on your server side, and will be able to carry out the next move with your information and to use it for their malicious purposes.

Keywords – Weka tool, Deep Learning, Neural Network

I. INTRODUCTION

Phishing is defined as a reverberation of a notable company's website that captures personal information from customers, such as usernames, passwords, and structured savings numbers. Mail spammers can be classified based on who they are trying to reach. Some telemarketers are spammers who send a few hundred or a big number of spontaneous e-mail messages to customers. Spammers are classified as follows: they continue to send messages at random but aren't really enthusiastic. They frequently spam or push resources that are irrelevant to the issue. Sees, knowledgeable news, and words regarding meetings are some of the examples. Phishing is not a new concept, but criminals, or phishers, have increasingly utilized it in recent years to steal personal information and commit economic and social crimes. In the last four to five years, the number of phishing assaults has increased dramatically. Phishing is a common practise that is simple to carry out at your destination.

The new term 'fishing' refers to an attacker's invitation to visit a fake site by imitating a website's appearance in order to obtain personal information from users such as usernames, passwords, financial information, account details, national security identifiers, and so on. Phishing is a new phrase coined from the word 'fishing.' The data gathered is utilized for prospective target advertisements or potentially identity theft and attacks (such as money transfers from one's account). Sending e-mails, messages that can lead to data theft or personal information, is a common attack strategy.

II. RELATED WORK

Financial institutions of all sizes are being targeted by clever, well-organized, and well-funded cyber criminals who are targeting commercial and retail account holders. All account holders are automatically protected against all sorts of fraud attacks with minimal disruption to legal online banking activity thanks to the availability of anomaly detection technologies that can be implemented rapidly and immediately. In addition to meeting FFIEC expectations, implementing anomaly detection will lower the total cost of fraud while increasing consumer loyalty and trust, according to the FFIEC.

As a result of the extremely effective data collection, the attributes of the suggested algorithm will be reduced to a bare minimum, which will improve the efficiency of classification systems. The URL of the permitted URL should be inserted into the classifier in order to determine whether the URL is legitimate or malicious. The classification of URLs is based on a newly developed fuzzy logical technique to particle filtering, which is based on fuzzy logic. The function extraction method is responsible for extracting certain characteristics. The proposed selection technique, which is based on the Rough Set Theory algorithm and the Gray Wolf Optimizer, is used to find the best URL functions for a given URL. When it comes to precision, the proposed method is logical (it can achieve up to 70% accuracy), and it simply necessitates an analysis of the features of URLs. As a web filter or a risk scaler, this model can be used during preprocessing to determine whether or not input URLs are friendly, or to estimate whether or not an input URL is harmful.

III.ARCHITECTURAL DESIGN

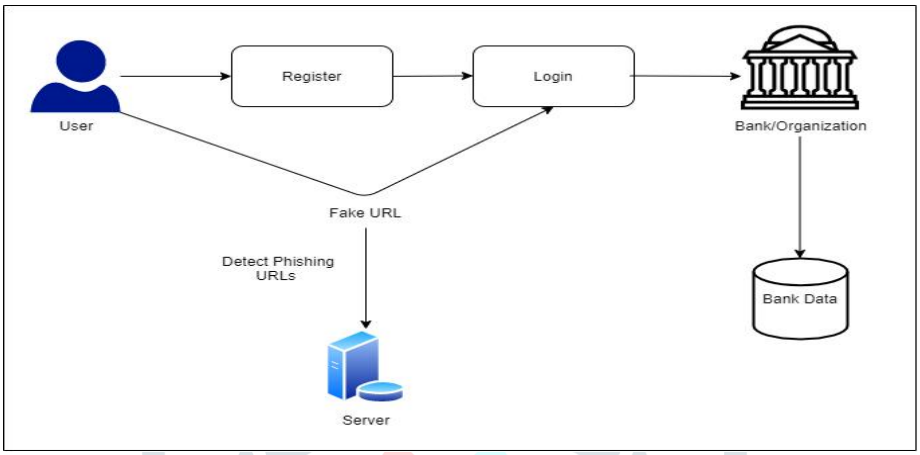


Fig. System Architecture

#Technology Used :-

1] Algorithms:-

Recurrent Neural Network(RNN)

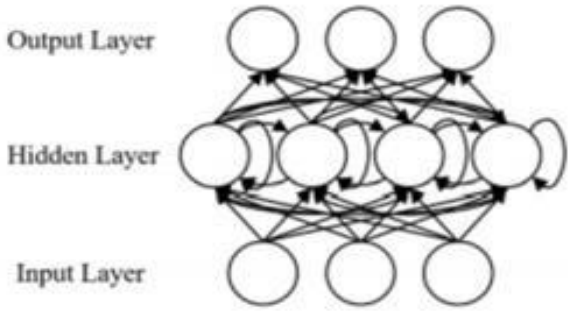


Fig. Recurrent Neural Network

As shown in Fig. , for a RNN, let our input x be a sequence whose length is T, x = x1 , x2 ,...,xT , and each item xt is a feature vector. At time step t, given the previous hidden layer state ht1 , the current hidden layer state ht and the output layer state yt can be calculated by,

ht = oh(whxt +Uhht −1 + bh)

yt = oy(wyht + by)

where Wh and Wy denote the input-to-hidden and hidden-to-output weight matrices, respectively, Uh is the matrix of the recurrent weights between the hidden layer and itself at two adjacent time steps, bh and by are the biases, and h and y denote the activation functions.

At each time step, the input is propagated in a standard feed forward fashion, and then, a learning rule is applied. The back connections lead to the result that the con-text units always maintain a copy of the previous values of the hidden units (since they propagate over the connections before the learning rule is applied). Thus, the network can maintain a state, allowing it to perform such tasks as sequence predic-tion that are beyond the power of standard multilayer perception.

Formula for calculating current state:

ht = (ht −1, xt)

where,

ht=current state

ht-1=Previous state

xt= Input state

Formula for applying Activation function:

ht = activation(whhht −1 + wxhxt)

where,

whh= Weight at recurrent neuron

wxh= Weight at input neuron

Formula for calculating output:

yt = whyht

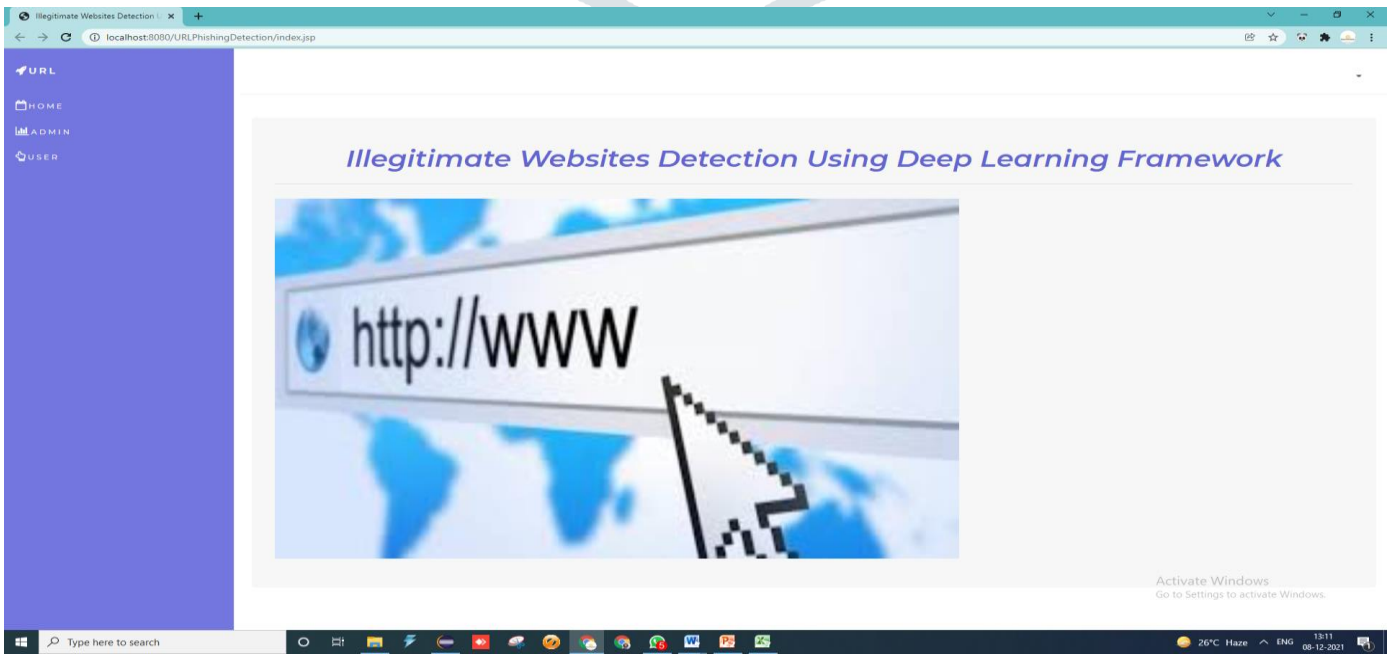
where,

yt=Output

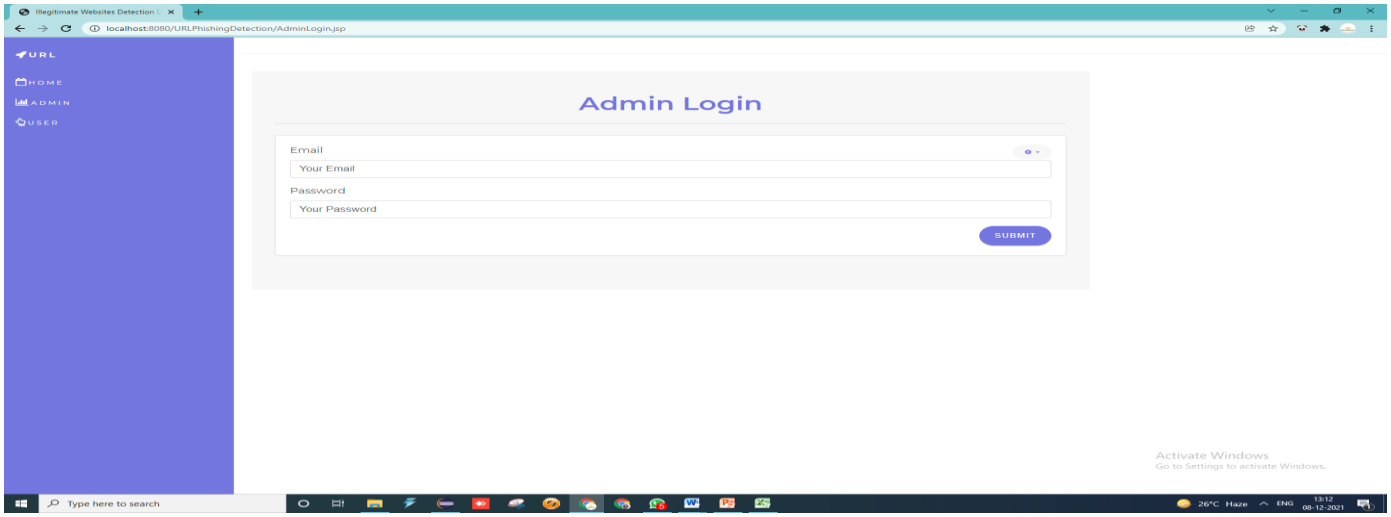
why=Weight at output layer

IV. EXPERIMENTAL RESULTS

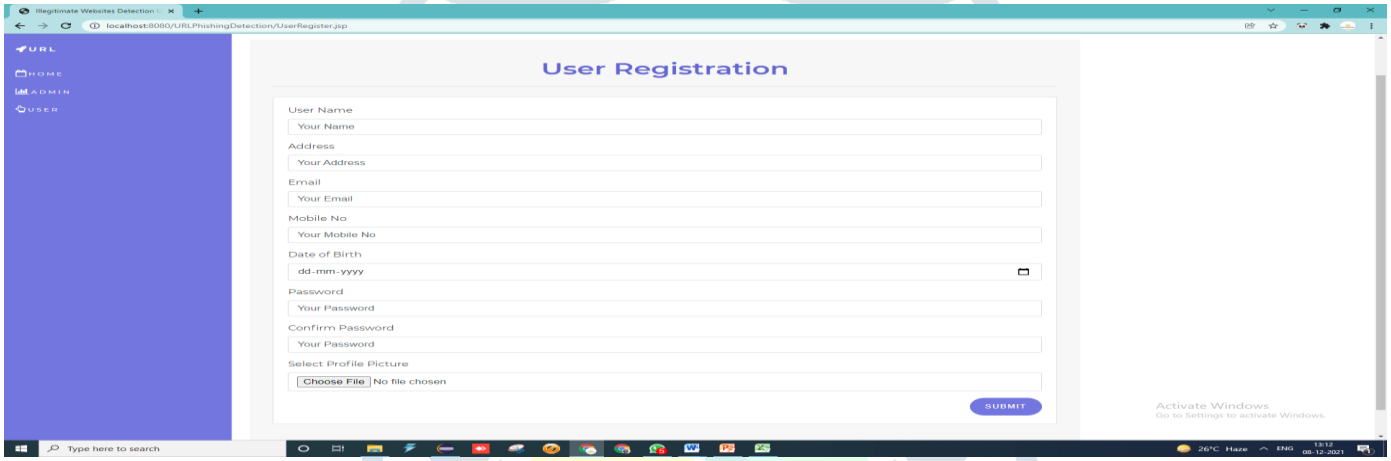
The fundamental principle behind the development of such a system is to ensure that financial information for a customer is safe, and so banks and other financial institutions provide various security measures to minimise the risk of unauthorized access to their online account. Online banking has been completely relayed on online transactions through various applications nowadays, so it is most important that this online banking activity is secured.



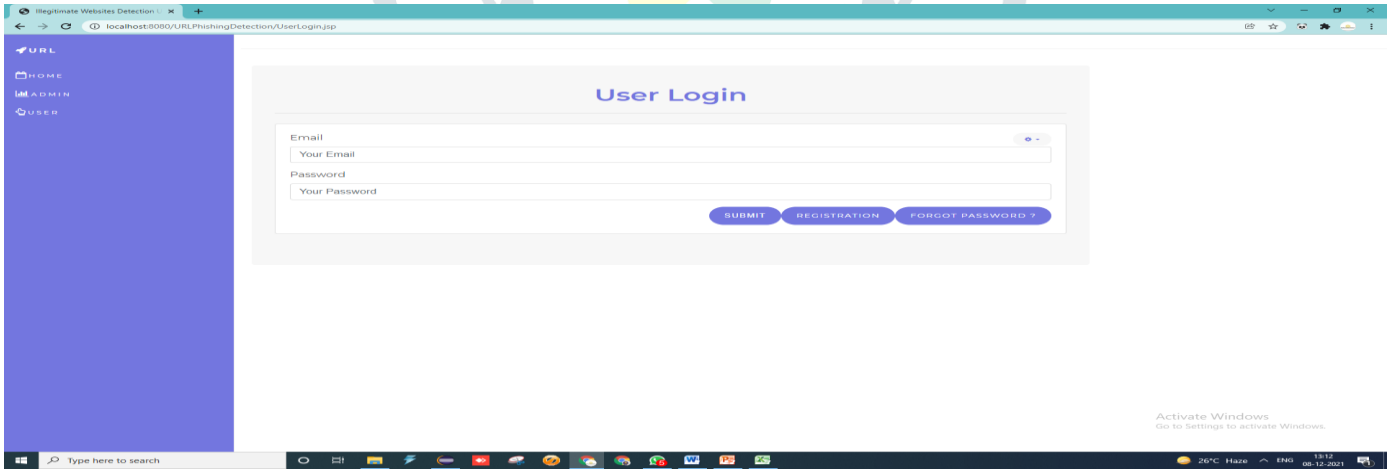
(a) Home Page



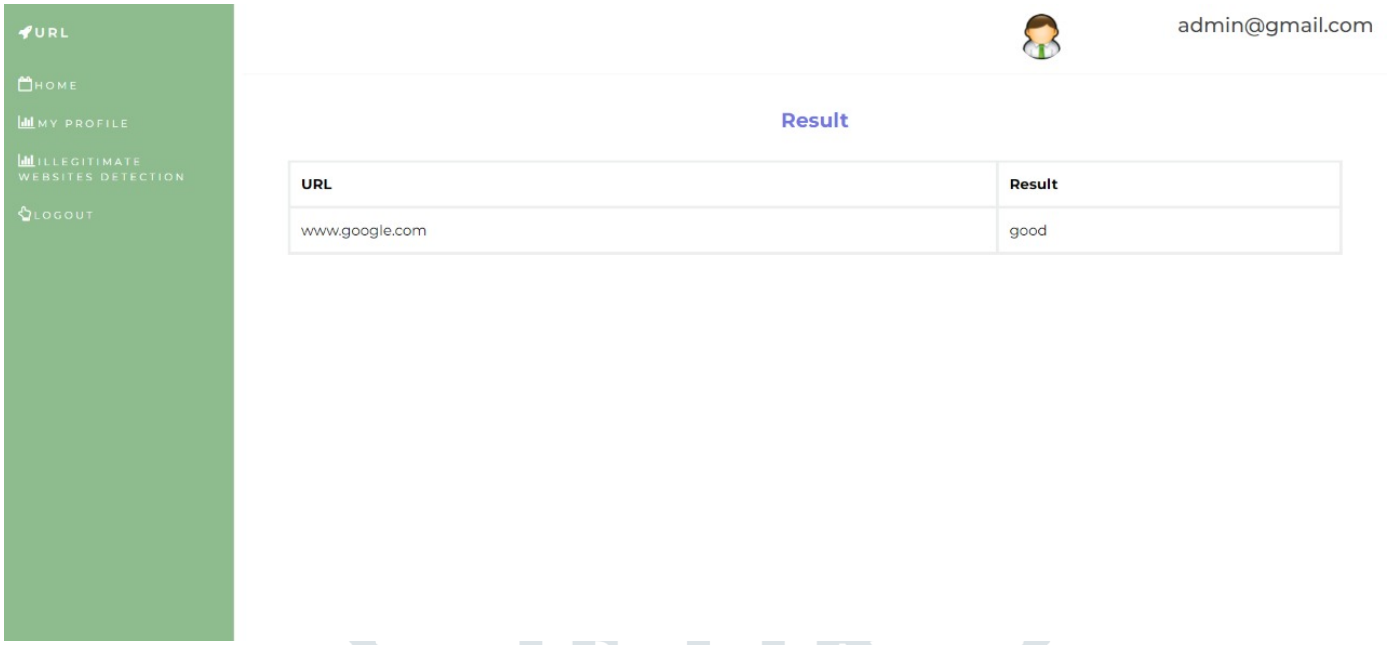
(b) Admin Login Page



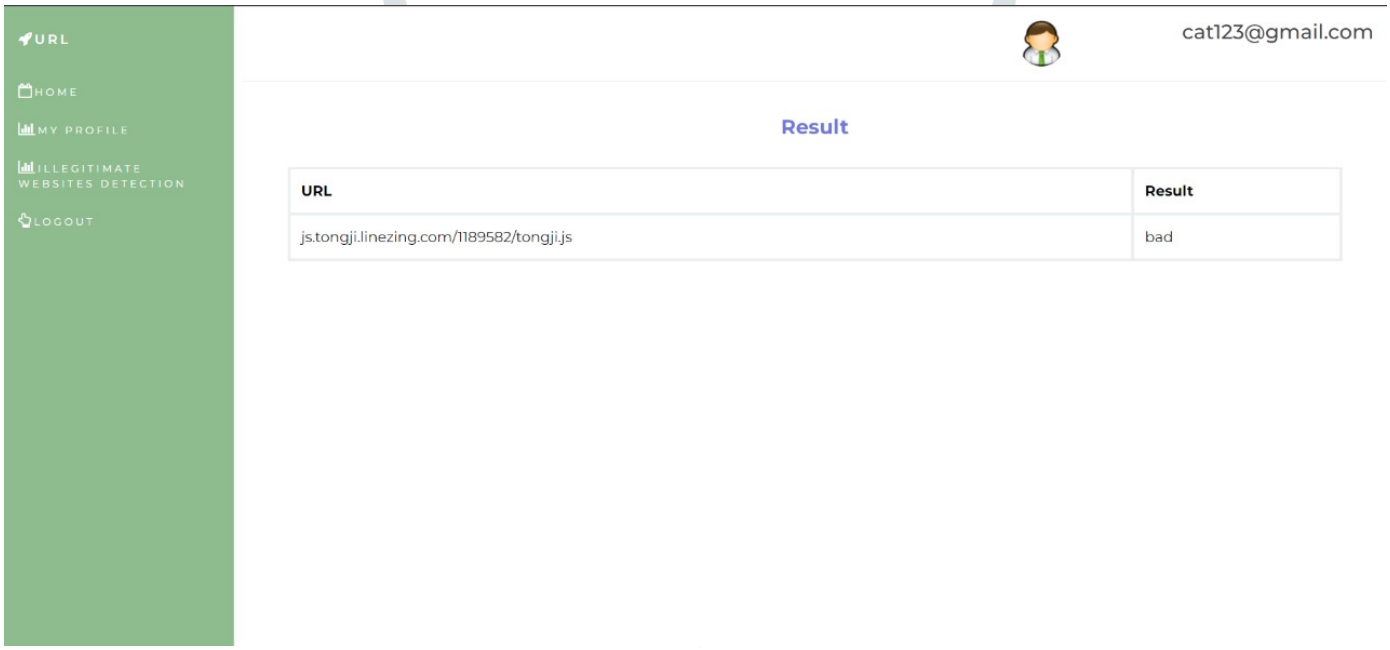
(c) User Registration Page



(d) User Login Page



(e) Result Page I



(f) Result Page II

V. CONCLUSION

Phishing is one of the most damaging web security threats. We have created a pre-diction model for the detection of Phishing websites by analyzing the attributes of the attack according to our study. The deep-seated learning model of the Deep re-current neural Network overcomes other machine learning models via prediction and achieves the highest precision.

VI. REFERENCES

1. Surbhi Gupta et al., "A Literature Survey on Social Engineering Attacks: Phish-ing Attack," in International Conference on Computing, Communi- cation and Automation (ICCCA2016), 2017, pp. 537-540.

2. Jian Mao, Wenqian Tian, Pei Li, Tao Wei, Zhenkai Liang, "Phishing- Alarm: Robust and Efficient Phishing Detection via Page Component Similarity".

3. Zou Futai, Gang Yuxiang, Pei Bei, Pan Li, Li Linsen, "Web Phishing De-tection Based on Graph Mining", Guardian Analytics,"A Practical Guide to Anomaly Detection Implications of meeting new FFIEC minimum expectations for layered security". Accessed: 08 Jan 2018.

4. Ibrahim Waziri Jr., "Website Forgery: Understanding Phishing Attacks Non-technical Countermeasures," in IEEE 2nd International Conference on Cyber Security and Cloud Computing, 2015,IEEE.

5. LongfeiWu et al,"Effective Defense Schemes for Phishing Attacks on Mobile Computing Platforms," IEEE 2016, pp. 6678-6691.

6. K. Rajitha and D. Vijayalakshmi, "Suspicious urls filtering using optimal rt-pfl: A novel feature selection based web url detection," in Smart Computing and Informatics, S. C. Satapathy, V. Bhateja, and S. Das, Eds. Singapore: Springer Singapore, 2018, pp. 227–235.

7. S. Kim, J. Kim, and B. B. Kang, "Malicious url protection based on attackers' habitual behavioral analysis," Computers Security, vol. 77, pp. 790 – 806, 2018.

8. B. B. Gupta, N. A. G. Arachchilage, and K. E. Psannis, "Defending against phishing attacks: taxonomy of methods, current issues and future directions," Telecommunication Systems, vol. 67, no. 2, pp. 247–267, Feb 2018.

9. A. K. Jain and B. B. Gupta, "A novel approach to protect against phishing attacks at client side using auto-updated white-list," EURASIP Journal on In-formation Security, vol. 2016, no. 1, p. 9, May 2016.

10. F. D. Abdi and L. Wenjuan, "Malicious url detection using convolutional neu-ral network," Journal International Journal of Computer Science, Engineering and Information Technology, vol. 7, no. 6, pp. 1–8, 2017.

11. E. Buber, B. Dırı, and O. K. Sahingoz, "Detecting phishing attacks from url by using nlp techniques," in 2017 International Conference on Computer Science and Engineering (UBMK), Oct 2017, pp. 337–342.

12. T. Shibahara, K. Yamanishi, Y. Takata, D. Chiba, M. Akiyama, T. Yagi, Y. Ohsita, and M. Murata, "Malicious url sequence detection using event de-noising convolutional neural network," in 2017 IEEE International Conference on Communications (ICC). IEEE, 2017, pp. 1–7.

13. M. Babagoli, M. P. Aghababa, and V. Solouk, "Heuristic nonlinear regression strategy for detecting phishing websites," Soft Computing, Feb 2018.

14. T. Peng, I. Harris, and Y. Sawa, "Detecting phishing attacks using natural language processing and machine learning," in 2018 IEEE 12th International Conference on Semantic Computing (ICSC), Jan 2018, pp. 300–301.

15.
O. K. Sahingoz, E. Buber, O. Demir, and B. Diri, "Machine learning based phishing detection from urls," Expert Systems with Applications, vol. 117, pp. 345–357, 2019.