YEAR	AUTHORS	OBJECTIVES	METHODOLOGY	LIMITATIONS
2006	Chandrasekaran.M	As an attack of social engineering, phishing email has caused tremendous financial loss to recipients. Therefore, there is an urgent need for phishing email detection with high accuracy. phishing emails detection based on Structural features. By analysing the emailheader structure, email-URL information, emailscript function and email psychological features, Experiments are performed on a dataset consisting of 500 legitimate emails and 500 phishing emails. The proposed approach achieved overall true-positive rate of 99%, false-positive rate of 9%, precision of 91.7% and accuracy of 95.00%. Furthermore, The results showed that psychological features can improve the accuracy of detection and reduce the false-positive rate.	The research aims to separate phishing emails message from regular messages, with the goal that the recipient is not affected by the phishing email on time. Phishing messages often contain specific words, whereby the recipient immediately performs specific malicious actions and leads to phishing. In this paper, we consider phishing detection a classification problem. Therefore, we require a classification algorithm and a feature set for a classification problem. We have raw emails message as input and in the training phase, a phishing label or benign label is assigned to each email. The proposed work begins by investigating the accuracy of phishing email detection using a group of features (header features, hyperlink features, and Complete features). Then, minimize the time and space required to extract and use these features by choosing the best features Next, the performance of the MLP and RF classification algorithms is examined on the set of extracted features and selected features by the IG method. Finally, the performance of the two classifiers is compared to determine the best phishing detection algorithm.	Time consuming, Huge number of features ,consuming memory.Non standard classifier . Time consuming because this technique has many layers to make the final result . Huge number of features . Many algorithm for classification which mean time consuming . Higher cost . Need large mail server and high memory requirement .Less accuracy because it depend on unsupervised learning,need feed continuously

2017	Hassan Y A Abutair	Many classifications techniques have been used and devised to combat phishing threats, but none of them is able to efficiently identify web phishing attacks due to the continuous change and the short life cycle of phishing websites. In this paper, we introduce a Case-Based Reasoning (CBR) Phishing Detection System (CBR-PDS). It mainly depends on CBR methodology as a core part. The proposed system is highly adaptive and dynamic as it can easily adapt to detect new phishing attacks with a relatively small data set in contrast to other classifiers that need to be heavily trained in advance. We test our system using different scenarios on a balanced 572 phishing and legitimate URLs. Experiments show that the CBR-PDS system accuracy exceeds 95.62%, yet it significantly enhances the classification accuracy with a small set of features and limited data sets.	Case based methodology is based on using historical data such as cases or experiences to predict a solution to the current problem. There are no typical cases or experiences, but there are similar ones. For that, adaptation should be considered in order to select the most similar case as a candidate solution and adding it as a revised one to the historical data to be used for future in predictions. CBR methodology is similar to Human thinking , so it can be used to solve different types of problems. A case as defined in is an 'experience of a solved problem', and can be represented in different ways. The word reasoning means a conclusion is drawn based on the registered cases within the case base. This conclusion forms the solution to the current problem, and may need further modifications or adaptations in order to withstand as a suitable solution. A case represents an experience to be used to derive a similar or close solution to the current problem. The main feature of CBR is the ability to emulate human thinking for solving problems from past experiences or cases as similar problems have similar solutions. Solving the current problem or predicting the state of the target URL against local case database follows the cycle depicted . In addition to the local case base, new experiences or cases should be acquired online frequently from different sources to maintain a current cases base to insure and maintain a high accuracy	An adaptive mobile phishing detection model based on a variation of input feature patterns using a case-based reasoning (CBR) technique is proposed in this work. An experimental analysis is conducted to demonstrate the design decision of our model and to verify the performance of our proposed model in handling the concept drift of mobile phishing attacks. The proposed model is evaluated with a large feature set that contains 1,065 features from 10 feature groups which are frequently collected from Android apps. Moreover, 5 cases of randomly combined patterns of features are created in order to provide a diversity of unknown patterns to mimic new real-world mobile apps. Six classification algorithms are chosen from different categories for the coverage usage of all classification nature on the diversion of feature sets
2021	Ashit kumar dutta	In recent years, advancements in Internet and cloud technologies have led to a significant increase in electronic trading in which consumers make online purchases and transactions. This growth leads to unauthorized access to users' sensitive information and	RQ3 stated that how ML method can be employed to identify a malicious or legitimate URL. To present a solution, authors proposed a framework as shown in Fig 3 for classifying URLs and identify the phishing URLs. Let Pm n¼0 xn be the set of URLs where m is the maximum limit for the number (n) of URLs. Let M, L 2 xn be the malicious and legitimate, accordingly. Suppose M and L contains the properties Pm and Pl, respectively. The proposed	Both SVM and NB are slow learners and does not store the previous results in the memory. Thus, the efficiency of the URL detector may be reduced. They compared the performance of different types of ML methods. However, there were no discussions about the retrieval capacity of the algorithms.

	damages the resources of an enterprise. Phishing is one of the familiar attacks that trick users to access malicious content and gain their information. In terms of website interface and uniform resource locator (URL), most phishing webpages look identical to the actual webpages. Various strategies for detecting phishing websites, such as blacklist, heuristic, Etc., have been suggested. However, due to inefficient security technologies, there is an exponential increase in the number of victims. The anonymous and uncontrollable framework of the Internet is more vulnerable to phishing attacks. Existing research works show that the performance of the phishing detection system is limited. There is a demand for an intelligent technique to protect users from the cyberattacks. In this study, the author proposed a URL detection technique based on machine learning approaches. A recurrent neural network method is employed to detect phishing URL. Researcher evaluated the proposed method with 7900 malicious and 5800 legitimate sites, respectively. The experiments' outcome shows that the proposed	framework employs RNN—LSTM to identify the properties Pm and Pl in an order to declare an URL as malicious or legitimate. The following equations from 1 to 4 presents the method for identifying the malicious URL. The term "recurring neural network" implies two broad groups of networks of a similar general structure, where one is a finite, and the other is an infinite input. Both network groups contains time dynamic behaviour. A recurrent network of finite input is a directed acyclic graph that can be replaced by a purely feedforward neural network, whereas a recurrent network of infinite input is a directed cyclical graph that cannot be modified. The modified version of RNN is LSTM. It is a deep learning method, which prevents the gradient problem of RNN. Multiple gates are employed for improving the performance of LSTM. In comparison with RNN, LSTM prevents back propagation. Each input of LSTM generates an output that becomes an input for the following layer or module of LSTM. Eqs 1 to 4 illustrates the concept of the proposed	The outcome of the experiments demonstrated that the performance of the system was better rather than other ML methods. However, It lacks in handling larger volume of data. Deep learning methods demand more time to produce an output. In addition, it processes the URL and matches with library to generate an output. The performance evaluation was based on crawler-based dataset. Thus, there is no assurance for the effectiveness of the URL detector with real time URLs. Authors employed an older dataset which can reduce the performance of the detector with real—time URLs. The performance of GA based URL detector was better; nonetheless, the predicting time was huge with complex set of URLs. The method employed a server for updating the page attributes that reduces the performance of the detecting system. The existing research shows that the performance of CNN is better for retrieving images rather than text. Neural Network based detection system can identify the impression of an adverse network by learning the environment.