LITERATURE SURVEY

S.NO	TITLE OF THE PAPER	METHODOLOGY USED	OBSERVATIONS	RESULTS + CONCLUSIONS	LIMITATIONS
1.	Phishing Website Detection Based on Machine Learning Algorithm (2020)	Analysed the features of the URL of the phishing website, used four machine learning algorithms for training.	Compares the similarity of snapshot of webpage with regular webpage	Similarity recommendatio nfor regular website corresponding to phishing website	Incase the number of features for each data point exceeds the number of training data samples , SVM will underperorm
2.	Detection of Phishing Emails using Machine Learning and Deep Learning (2022)	Machine learning and deep learning techniques on an imbalanced dataset	Proposed model is deployed through the web application FLASK PYTHON	Detect the email by giving information to the user about whether this email is genuine or fraudulent.	More potential for security risks, slower MVP development in most cases.
3.	Phishing Attacks Detection using Machine Learning Approach (2020)	Random forest and decision tree algorithms are used in dataset of phishing attacks	The attributes are analysed using principal component analysis (PCA)	A maximum accuracy of 97% was achieved through the random forest algorithm.	It needs a quality datasets. Noisy data and outliers have to be avoided.
4.	URL – based Phishing Websites Detection via Machine Learning (2021)	Neural networks and decision trees are employed to learn data patterns in websites URLs	System is evaluated on a recent phishing websites dataset using classification accuracy as a performance indicator	Decision trees models provide 97.40% classification accuracy on the almost balanced-class dataset.	Does not work well with high dimensionality as this will complicate the distance calculating process to calculate distance or each dimension.

5.	A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques (2022)	The most frequently used NLP techniques are found to be TF-IDF and word embeddings. Furthermore, the most commonly used datasets for benchmarking phishing email detection methods is the Nazario phishing corpus	It is unique in the sense that it relates works to their openly available tools and resources.	Helpful for the scientific community, especially in the field of NLP application in cybersecurity problems	The analysis of the presented works revealed that not much work had been performed on Arabic language phishing emails using NLP techniques. Therefore, many open issues are associated with Arabic phishing email detection.
6.	Identification and Analysis of Phishing Website based on Machine Learning Methods (2022)	Analyses and compare phishing website and legitimate by analyzing the data collected from open-source platforms and proposed a method to detect fake sites using Decision Tree and Random Forest approaches	Supports the purpose of identifying the best phishing website detection where Decision Tree and Random Forest were trained, tested and achieved high number of feature importance detection and accuracy rate	The result demonstrates that Random Forest has the best performance in phishing website detection compared to Decision Tree.	State-of-the-art comparison is missing and achieved accuracy is not reported expicitly.

7.	Feature Selection for Machine Learning – Based Phising Websites Detection (2021)	Combined two datasets with 30 and 48 features which identified13 optimal features for a more robust model	The best models built on all features using the random forest algorithm scored lower on the 30 features dataset, and achieved better performance on the 48 features dataset	The best model on the 13 features achieved an accuracy of 0.937	Assumption of linearity between dependent variable and independent variables is improper.
8.	Towards the Detection of Phishing Attacks (2020)	List-based, machine learning, visual similarity, Heuristic-based approach are used	A hybrid approach of phishing attack detection	Efficiently detect legitimate websites easily	Sometimes, the method can provide an inaccurate solution or judgment about how commonly things appear
9.	Detecting Phishing Websites Using Machine Learning (2022)	Data consisting of 86 features and 11,430 total URLs, are trained with 8 ML algorithms.	Accuracy of 8 different ML algorithms are compared with each other	The highest accuracy of 96.6 using XG Boost .	Does not work well with high dimensionality as this will complicate the distance calculating process to calculate distance or each dimension.

10.	Phishing Attacks Detection using Machine Learning and Deep Learning Models (2022)	Two separate datasets and highest correlated features are used which comprised of a combination of content-based, URL lexical-based, and domain-based features and then ML models were applied	The results also aimed to identify the best features that influence the model in identifying phishing websites	Random Forest (RF) algorithm achieved the highest accuracy for both datasets.	This algorithm is not applicable for applications in classification issues.
11.	Securellot Environment : Federated Learning Empowered Approach for Securing IIoT From Data Breach (2022)	A Safe data sharing architecture for various IIoT devices using Federated Learning is proposed which incorporates FL into the edge computing consensus process	High efficiency, and better security, according to numerical findings generated by experimenting deep learning models	Framework named as SecureIIoT is achieved 99.79% accuracy by detecting attacks as a binary classification problem.	Federated learning requires frequent communicatio n between nodes during the learning process.
12.	A lightweight and proactive rule-based incremental construction approach to detect phishing scam (2022)	A lightweight and proactive rule-based incremental construction approach to detect any unknown phishing URLs	This application can detect the zero-day and spear phishing attacks with a detection rate of 89.12% and 76.2%, respectively	The application shows the precision level higher than the previous model developed and other phishing techniques.	Using proactive strategies could be the potential risk of creating a disruption where none previously was

13.	Detection of Malicious Cyber Fraud using Machine Learning Techniqures (2022)	Ascertain malicious URLs is formulated on RF,SVM,CNN,DNN .	Model is trained to detect URL behaviour and attributes.	Based on URL behaviour and attributes,the malicious URL is detected efficiently	Variable selection is not explicitly mentioned and state-of-the-art comparison is missing.
14.	Phishing Websites Detection using Machine Learning with URL Analysis (2022)	Uses URLs as a dataset containing 6000 URLs to detect phishing websites.	Performance analysis of 8 ML algorithms provides the accuracy which is being compared	Multilayer perceptron algorithm got the highest accuracy of 85.41%	Does not work well with large dataset as calculating distances between each data instance would be very costly.
15.	Survey on Detection and Prevention of Phishing Websites using Machine Learning (2021)	The task to execute the frameworks with good efficiency, exactness, and cost-effectively is done using the four classification models are KNN, Kernel-SVM, Random Forest Classifier and Decision tree	With machine learning, cybersecurity systems can analyze patterns and learn from them to assist prevent similar attacks and answer changing behavior.	The four classification models were discussed and analysed in term of the merits, demerits and performance	Does not work well with high dimensionality as this will complicate the distance calculating process to calculate distance or each dimension.