International Journal of
Electrical and Computer System Design
ISSN: 2582-8134 www.ijecsd.com

COLLABORATIVE FILTERING OF MALICIOUS INFORMATION FROM THE MULTIMEDIA DATA USING DEEP BELIEF NEURAL NETWORK

Ms Gomathy M *[1], Dr.A.Vidhya [2]

*1 Research Scholar, School of Computing Sciences, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, Tamil Nadu, India.
marimuthu.gomathy@gmail.com.

2 Assistant Professor, Department of Information Technology, School of Computing Sciences, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai., Tamil Nadu, India.
avidhya.scs@velsuniv.ac.in

Abstract: Every aspect of daily life has been impacted by the Internet's debut, which allowed for global virtual collaborations. Owing to the internet's wide distribution, the exponential increase of mobile data, and the widespread use of online forums, online crime has become more common. Online users, including but not limited to service providers, face a serious threat from unwanted information since it harms their reputations and services. Consequently, creating an intelligent model to filter unsolicited information is required. Unwanted information is categorized and filtered using contemporary machine learning algorithms. While processing all the text, image, and video files, some standard procedures are insecurely created, making it difficult to detect malware in the files. Due to the exponential growth in the creation of new ransomware, malware identification has become a challenging research topic. Businesses and commoners find it difficult to protect themselves against malware in the digital environment, which emphasizes the importance of developing efficient spyware protection techniques. Considering the above context, the objective of this systematic review is two-fold, (1) To examine and comprehend the numerous problems caused by malicious programs in virtual platforms and (2) to assess and suggest a method using Deep Neural Network for classifying a sample as benign or malicious with high accuracy and minimal overhead.

Keywords : Adolescents; Cybercrime; Deep Neural Networks; Malware; Victimization

1. Introduction

Advancements in technology allow people to be more connected than ever. Although there is evidence of the detrimental consequences of technology, other forms of technology may have contributed to good improvements in the world. Digital platforms suffer a great loss in terms of their clients and their reputations on receiving irrelevant data contained within text, image, and video files [1]. The exponential increase in the number of unsolicited messages has influenced to develop screening strategies that are more dependable and efficient.

Technology had a significant impact on exploring information as well as the exchange of information in digital platforms. As enormous data have been transmitted in cyber platforms, there is a rapid growth in the volume of unsolicited and unwanted messages floated in them. Moreover, unsolicited information poses a major threat to the information security system.

Proceeding of "Technology Integration for Sustainable Development: An International E-Conference on Electrical, Electronics, Computer Science and Mechanical Engineering. (EECM-2023)". Organized by SJUIT.

Antivirus software is a crucial barrier against malware assaults. Usually, a signature-based methodology is applied for malware detection. However, the above approach does not provide security against zero-day attacks. Additionally, by utilizing various encryption approaches, malware-generating resources like Zeus [1] might produce hundreds of different versions of identical malware. Therefore, we propose a deep neural network-based malware analysis system.

In this paper, we surveyed and divided the extant material into two categories of study. 1)feature extraction: Static analysis and dynamic analysis are two distinct methods for generating characteristics in malware analysis. Unlike static analysis, dynamic analysis is a malware analysis technique that involves the execution of the sample, and then studying the sample's behavior as it is being run. The dynamic analysis allows one to bypass obfuscation by observing the behavior of the samples rather than trying to decipher the sample's contents and behaviors. Moreover, Malware is intrinsically unsafe and researchers use cuckoo sandbox to extract the features for constructing the malware dataset. Ye et al. [02] used Windows API calls and showed that Open Process, CloseHandle, and CopyFileA few Application Programming Interface (API) class usually occurs along with malicious executables. Byte-level n-gram can be utilized to gather more information on malicious content from the code as compared to portable executable headers. Additionally, few opcode strings carry crucial semantic information that makes it possible to quickly determine an attacker's intent. Different malware detection

algorithms have made use of a number of feature selection techniques, including document frequency [8], information gain [7], and max-relevance algorithm [3]. Auto-encoders have been utilized to reduce the size of the memory when dealing with large voluminous datasets. (2) Building Classification Models: After feature extraction, the feature vectors are used by the classification algorithms to build a model detection of malware in the data set. Deep neural networks have the ability to spot hidden relationships and patterns in raw data. It can be utilized to cluster and classify the dataset and can continuously learn and improve. In this paper, a collaborative modal of deep learning algorithms is utilized to filter the malicious content from the data.

The main contribution of the work involves the following:

- The author discusses the problems in the virtual platform on using Malicious Data.
- The authors detect the presence of malicious data from the text and multimedia data.
- The authors develop a deep belief neural network for the mitigation of unsolicited data.

People can use neural networks to address complex problems in real-world environments. Input-output interactions that are complicated and nonlinear can be learned and modeled by them. Predictions and generalizations can be drawn from this in order to uncover latent associations, patterns, and forecasts. Neural networks can thereby enhance decision-making in fields like malware detection.

Problems in Virtual Platforms:

In this contemporary era, with the aid of the internet, communication has spread across a number of online communication platforms. Twitter, Facebook, and other internet media services have been integrated into our daily routine for private and professional interactions in addition to electronic mail and instant messaging. Most organization encourages their employees to have virtual communication and update to the current digital trends. Since there is less face-to-face interaction and shared office space due to more flexibility, more information must be made available to colleagues online. With the help of cloud services and decentralized access, most of the file sharing and communication is conducted by third parties like social networks or other types of platforms. Most people communicate extremely sensitive information on virtual platforms without knowing or caring about security and privacy, which encourages cybercrime.

Nowadays, a significant number of individuals use online social networks for communication. Any person can share more information with anyone in an online platform. However, some social network users abuse the characteristics of these platforms and encourage the dissemination of harmful content. They accomplish this by uploading harmful files which spread rapidly. There is no reliable method for swiftly discovering and effectively removing these harmful files.

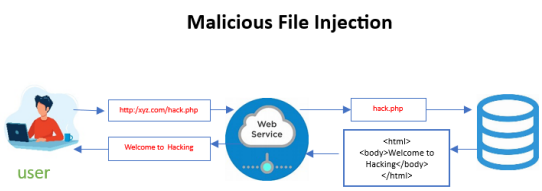

Fig 1 : Malicious Infection

Eventually, social media platforms struggle with their enormous user bases. This has prompted many people to utilize the network to commit cybercrime against other users. Inappropriate account activity is one type of cybercrime mostly used on social media platforms. Problematic accounts include spambots and phoney followers which can negatively impact fellow users. Unwanted messages from spambots are frequently sent to other users that can raise other accounts' following counts, which might indicate influence or reliability.

2. Literature Survey

The research takes a look at collaborative filtering algorithms, the challenges they face, and the robust training techniques that could be adopted from elsewhere in the text in this section. This portion will be located after the introduction.

There are two primary schools of thought when it comes to collaborative filtering. The first school of thought is the school of thought that depends on nearest neighbors and the second school of thought is the school of thought that depends on models [7]-[9]. To determine the inclinations of a single user or item, algorithms that are based on the concept of the nearest neighbor look for other users or other items that have profiles that are comparable to their own. Model-based collaborative filtering involves the construction and approximation of parametric models based on witnessed interactions between users and the products that they are interested in.

Among the model-based approaches, matrix factorization algorithms are the model-based joint filtering strategies that see the most widespread application. The technique that is used to generate predictions is called the multiplication of estimated user and object factors. These factors are obtained through the factorization of previous interactions that have taken place between users and objects [10]-[12]. An additional canonical school of thought makes use of contextual information to construct an inductive matrix completion model [13–16], which ultimately leads to enhanced joint filtering performance.

When creating a helpful recommendation system, it is essential to model the latent representations of the users as well as the products in an accurate manner. The effectiveness of neural networks in representation learning has contributed to the rise in popularity of a method known as Neural Network-Based Collaborative Filtering (NeuCF) in recent years. This method promises to deliver an outstanding performance across a variety of benchmarks and has become increasingly popular in recent times. Users and products are paired off with one another to produce features, which are then input into a combined filtering model that is built on neural networks [17].

The various layers of the neural network and the activation functions that are used on them combine to produce a latent space that includes features of both the user and the object. By combining the neural network with the activation functions, this domain is produced. These characteristics are imagined to exist within this region.

In [10], a bilinear decoder is used to transform the hidden person and object factors into ratings, whereas in reference number [9], a mathematical operation in linear classifier formation is used. In the following, a comparison will be made between these two approaches. Even though autoencoder-based methods [18] are considered to be a preferable case of performing standard proposed architecture, they still take inputs in the shape of items or useful features. This is the case even though autoencoder-based methods can be viewed as a standard Neural

Network-Based Collaborative Filtering (NeuCF) architecture.

Both [11] and [19] implement a stochastic method that has an organizational structure that is analogous to that of autoencoder-based models. Stochastic units with a particular distribution, such as binary or Gaussian, are used in [19] in favor of ReLU or logistic units. On the other hand, [11] presents a model that is more complicated and can work with high-dimensional binary vectors.

The algorithms are subjected to a comprehensive research endeavor, even though it has been demonstrated that the algorithms are effective in improving the performance of recommendation systems. We are going to look at previous research on offensive and defensive strategies that we have accessible to us right now, and we will look at both types of strategies [20, 21].

Carry out some investigation on the difficulties that neighbor-based collaborative filtering is encountering in the modern world. An attack known as data poisoning attacks factorization-based recommendation systems while those systems are in the process of being trained through the joint filtering process. This method of attack, which was investigated in [22], is directed at factorization-based recommendation systems. It is clear from the evidence presented in [23] that both dependability and accuracy are present in equal proportion. Deep Neural Networks (DNN) [24] are currently the subject of intensive investigation due to the perceived deficiencies of these networks.

The discovery in [25] that the smoothness assumption that forms the basis of the foundation of many kernel methods does not hold implies that the classification results obtained by DNN might be very different if a minute, imperceptible perturbation is added to the image. This is because the smoothness assumption is at the foundation of many kernel methods. When it comes to the production of examples of aggressive behavior, various attacks assert that they have achieved extraordinary levels of success. These strategies are modifiable in a very straightforward manner and have the potential to be utilized productively in opposition to neural collaborative filtration algorithms.

It is a significantly more difficult task to defend against an adversarial attack than it is to launch one, and this is especially true in the field of recommendation systems where the stakes are particularly high. A network that is trained with soft labels is less likely to overfit the training data, which makes it more resistant to attacks from deliberately perturbed inputs that fall within the blind spots of non-linear networks. In other words, a network that is trained with soft labels is less likely to overfit the training data. A network that is trained with soft labels has a greater chance of producing accurate results. It has the potential to substantially mitigate the detrimental effects that adversarial occurrences have on DNNs, which is a significant possibility.

3. Problem Identification and Statements

3.1 Proposed Method

In the proposed study, three different models based on the objectives are designed that involve the following: the study develops a multi-modal approach to process text, image, and video data at the same instance of time that involves: pre-processing, feature extraction, and classification. Hybrid modeling using deep belief neural network that classifies both the text, image, and video data at the same instance of time. The model is designed in such a way that it operates on multi-modal output in providing better classification of instances than ever required by the other systems. Finally, the system validates if the classified content from the classifier is malicious or not.

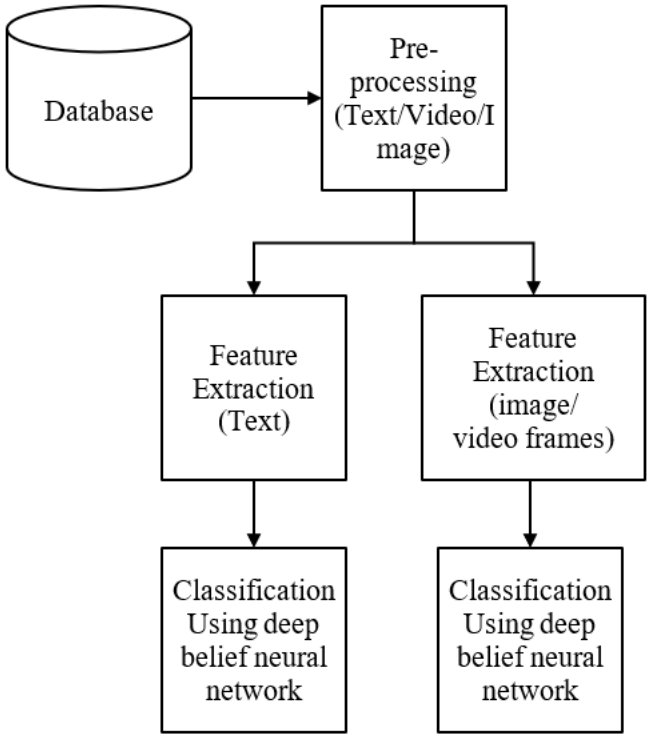

Fig 2: Proposed Framework

3.2 Preprocessing

The first and foremost stage performed in text mining is preprocessing. As the data contains noises and missing values, the preprocessing method helps to extract crucial features from the text. In this paper, we use the word2vec skip-gram model. In addition to identifying the word itself, a skip-gram model also predicts the word's context. The model is trained using skip-grams, which involve token skipping. A collection of skip-gram pairs (target word, context word) where the context word exists in the target word's nearby context can be used to indicate the context of a word. With this form of preprocessing, a Deep learning model can be used to improve the precision and effectiveness of a learning model.

3.3 Feature Extraction of Texts

An essential and significant task in text mining and information retrieval is the selection of text features. Handcrafted characteristics are essential for conventional approaches in determining the features. Deep learning techniques allow for the acquisition of new, efficient feature representations from training data as the traditional strategy is time-consuming. Deep learning automatically learns features from vast amounts of data, in contrast to conventional methods that rely on the prior experience of designers Big data with millions of parameters can be used by deep learning to autonomously acquire feature representation.

RNNs (Recurrent Neural Networks) are capable of capturing information while modeling text since they have a recollection of previous inputs. The context of the input

sequence's context, which consists of millions of factors, can be automatically learned via deep learning. As a result, they are helpful for projects like language modeling, where the context in which a word is used determines its meaning. The MultiLayer Perceptron, also known as MLP, is utilized to acquire a representation of the text from the words that are contained in the word vector that is fed into the Recurrent Neural Network (RNN). This is accomplished by feeding the word vector into the RNN. The following is an illustration of one model that could be used as a potential basis for the method. GRU (Gated Recurrent Unit) is used in standard Recurrent Neural Networks (RNN).

$e_i = W_e(w_i), i \in [1,n]$

$h_i^f = GRU^f(e_i, h_{i-1}^f)$

$h_i^b = GRU^b(e_{n-i}, h_{i-1}^b)$

$t = MLP\ 1/n(\sum_{i=1}^{n}(h_i^f + h_i^b)/2$

where the word embedding matrix id represented as

e_i denotes the ith embedded word.

$GRU^f(x)$ and $GRU^b(x)$ denotes the forward and the backward GRU.

h_i^f and h_i^b indicate the hidden state of the GRU^f and GRU^b at step i.

3.4 Feature Extraction of Image

One of the traditional methods to map between images and text is by using visual extraction and textual features extraction and applying a machine learning algorithm to learn the mapping between them. Once the model created by the traditional method is evaluated, it can be used to predict the relationship between new images and text by inputting their corresponding features into the trained model. However, with the advancements in deep learning, end-to-end models can be combined to learn visual and textual representations that have gained outperformed traditional techniques.

In feature extraction using Multi-View Sentiment Analysis (MVSA), the image is given as input to extract the features from it. The hierarchical features can be extracted at various levels of abstraction by using a Convolutional Neural Network (CNN) on the extracted features. Low-level regional features to high-level global features are all captured in various levels of detail by the feature maps. Recognize linkages between various spatial places on each feature map by paying close attention to each one separately. Based on their significance in portraying the features, it gives different locations weights. A fused attention map is created by combining the attention maps that were derived from various feature maps or layers. Both regional and global dependencies within the image are represented by this fused attention map. Multiply the original feature maps by the fused attention map, element by element. This operation emphasizes the important features based on the attention weights. Aggregation methods, such as spatial pyramid pooling, combine the weighted feature maps to create a fixed-length feature vector that incorporates all the required data from the image. The resulting feature vector can be used for the classification of unsolicited images from others. The fusion module that has been proposed takes the differentiation between global and local characteristics and generates dynamic weights to appropriately balance both sets of factors, as a result.

Classification using Deep Belief Neural Network

An unsupervised deep learning model called a Deep Belief Network (DBN) is one that has many layers of hidden units. To develop hierarchical data representations, it combines the ideas of restricted Boltzmann machines (RBMs) and deep learning.

A Restricted Boltzmann Machine (RBM) is a type of generative stochastic artificial neural network. It is a building block often used in the construction of Deep Belief Networks (DBNs) and other deep learning models. An RBM consists of two symmetrically connected layers: a visible layer and a concealed layer. While the layer that is concealed records the learned features, the visible layer displays the input data. The most common training algorithm for RBMs is called Contrastive Divergence (CD). CD approximates the gradient of the log-likelihood of the training data by performing a few steps of Gibbs sampling. CD training iteratively updates the weights and biases of the RBM based on variations in the probabilities for both the beneficial and detrimental phases.

It is composed of two levels, the first of which can be observed with the naked eye, and a second group of layers that are trickier to distinguish from one another. The task of distributing freshly arrived data to the many concealed layers, which oversee carrying out the machine learning process, falls under the purview of the visible layer.

Training the model using RBM uses Gibbs sampling and Contrastive Divergence Step. To predict the hidden vector from the input vector v, p(h| v) can be utilized or p(v| h) is used in Gibbs sampling.

$$p(v_i = 1 \mid h) = \frac{1}{1 + e^{(-(a_i + w_i h_j))}} = \tilde{O}(a_i + \sum h_j w_{ij})$$

After k iterations, we get a second input vector, v_k, which is made from the first input value, v_0.

$$p(h_i = 1 \mid v) = \frac{1}{1 + e^{(-(b_j + w_j v_i)}} = \sigma(b_j + \sum_i v_j w_{ij})$$

The weight matrix is updated throughout the contrastive divergence step. It makes use of the vectors v_0 and v_k to analyse the activation probability for the hidden variables h_0 and h_k.

$$p(v_i = 1 \mid h) = \frac{1}{1 + e^{(-(a_i + w_i h_j))}} = \tilde{O}(a_i + \sum h_j w_{ij}$$

Using the input vectors v_0 and v_k, the updated matrix is calculated by comparing the outer products of the probabilities.

$$\Delta W = v_0 \otimes P(h_0 \mid v_0) - v_k \otimes P(h_k \mid v_k) - v_k$$

We can use gradient descent to analyze new weight from the recently updated weighted matrix.

$$W_{new} = W_{old} + \Delta W$$

The proposed multimodal model uses the Adam algorithm to classify the multimedia information. The classification model helps to identify the classify the solicited information from the malware data set wither text or images. This model can be utilized in virtual platforms to restrict access to unsolicited information and combat the cybercrimes

4. Experimental Results and Discussions

In this paper, we utilized Kaggle dataset for text [26] and for Image dataset [27] is used. To evaluate the performance of the proposed model, various evaluation metrics were

considered. During the performance evaluation process, factors including accuracy, precision, memory, and f-score were taken into consideration.

The metrics are shown in the following equation:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN},$$

$$\text{precision} = \frac{TP}{TP + FP},$$

$$\text{recall rate} = \frac{TP}{TP + FN},$$

where TP is the number of true positive samples, FN is the number of false negative samples, FP is the number of false positive samples, and TN is the number of true negative samples. Accuracy is measured in terms of a percentage and denotes the proportion of the total number of occurrences (N) for which an accurate identification was made. The ratio of the predicted value to the actual value is the metric that is used to assess the accuracy of a positive prediction. This ratio is determined for a specific collection of data pertaining to favorable predictions.

In addition, a variety of hybrid model evaluations were carried out, with a total of one hundred and ten repetitions and a population of sixteen groups for each evaluation. The algorithm that was utilized was Adam, and the learning rate that was utilized for it was 0.001. The following figure shows that the proposed model outperforms the already-existing models

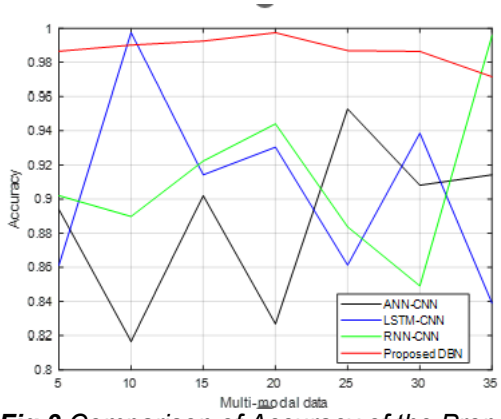

Fig.3 *Comparison of Accuracy of the Proposed Model*

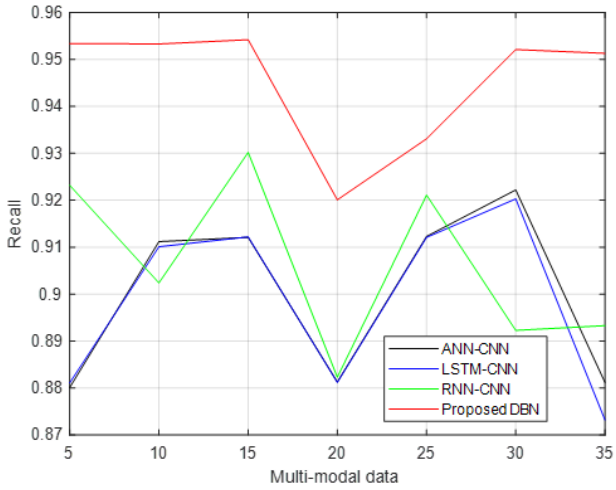

Fig.4 *Comparison of Recall of the Proposed Model*

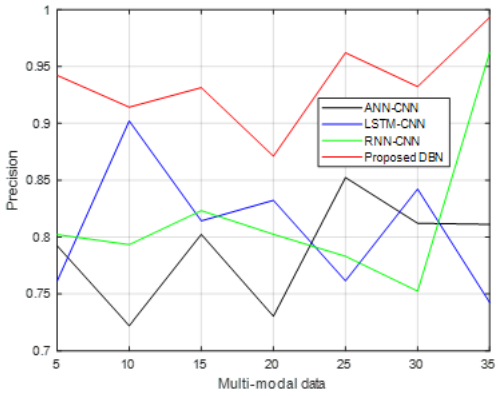

Fig.5 *Comparison of Precision of the Proposed Model*

5. Conclusion

The research that was recommended led to the creation of three distinct models, each of which was modified to accomplish one of the following three distinct goals: through pre-processing, feature extraction, and classification, a multi-modal approach for processing text, image, and video data simultaneously is developed as part of this research. Through this research, a hybrid model based on a deep belief neural network for the simultaneous classification of texts, photos, and videos was created. Compared to the earlier hybrid models that were already in use, the model is highly accurate at detecting and classifying malware data.

References:

1. Mishra, S., & Soni, D. (2020). Smishing Detector: A security model to detect smishing through SMS content analysis and URL behavior analysis. *Future Generation Computer Systems*, *108*, 803-815.
2. Jardine, E. (2019). Online content moderation and the Dark Web: Policy responses to radicalizing hate speech and malicious content on the Darknet. *First Monday*.
3. Alharthi, R., Alhothali, A., & Moria, K. (2021). A real-time deep-learning approach for filtering Arabic low-quality content and accounts on Twitter. *Information Systems*, *99*, 101740.
4. Islam, T., Latif, S., & Ahmed, N. (2019, May). Using social networks to detect malicious bangla text content. In *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)* (pp. 1-4). IEEE.
5. Sahoo, S. R., & Gupta, B. B. (2020). Classification of spammer and nonspammer content in online social network using genetic algorithm-based feature selection. *Enterprise Information Systems*, *14*(5), 710-736.
6. Crişan, A., Florea, G., Halasz, L., Lemnaru, C., & Oprisa, C. (2020, September). Detecting malicious URLs based on machine learning algorithms and word embeddings. In *2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP)* (pp. 187-193). IEEE.
7. Makkar, A., Kumar, N., Zomaya, A. Y., & Dhiman, S. (2020). SPAMI: A cognitive spam protector for advertisement malicious images. *Information Sciences*, *540*, 17-37.
8. Kang, D., Li, X., Stoica, I., Guestrin, C., Zaharia, M., & Hashimoto, T. (2023). Exploiting Programmatic Behavior of LLMs: Dual-Use Through Standard Security Attacks. *arXiv preprint arXiv:2302.05733*.
9. Begum, A., & Badugu, S. (2020). A study of malicious url detection using machine learning and heuristic approaches. In *Advances in Decision Sciences, Image Processing, Security and Computer Vision: International*

Conference on Emerging Trends in Engineering (ICETE), Vol. 2 (pp. 587-597). Springer International Publishing.
10. He, X., Gong, Q., Chen, Y., Zhang, Y., Wang, X., & Fu, X. (2021). DatingSec: Detecting malicious accounts in dating apps using a content-based attention network. *IEEE Transactions on Dependable and Secure Computing*, *18*(5), 2193-2208.
11. Naveen, I. N. V. D., Manamohana, K., & Verma, R. (2019). Detection of malicious URLs using machine learning techniques. *International Journal of Innovative Technology and Exploring Engineering*, *8*(4S2), 389-393.
12. Fang, Y., Xu, Y., Jia, P., & Huang, C. (2020). Providing email privacy by preventing webmail from loading malicious XSS payloads. *Applied Sciences*, *10*(13), 4425.
13. Afreen, A., Aslam, M., & Ahmed, S. (2020, October). Analysis of fileless malware and its evasive behavior. In *2020 International Conference on Cyber Warfare and Security (ICCWS)* (pp. 1-8). IEEE.
14. Sahoo, S. R., & Gupta, B. B. (2019). Hybrid approach for detection of malicious profiles in twitter. *Computers & Electrical Engineering*, *76*, 65-81.
15. Alshamrani, S., Abusnaina, A., & Mohaisen, D. (2020, November). Hiding in plain sight: A measurement and analysis of kids' exposure to malicious urls on youtube. In *2020 IEEE/ACM Symposium on Edge Computing (SEC)* (pp. 321-326). IEEE.
16. Collins, B., Hoang, D. T., Nguyen, N. T., & Hwang, D. (2021). Trends in combating fake news on social media–a survey. *Journal of Information and Telecommunication*, *5*(2), 247-266.
17. Yousaf, K., & Nawaz, T. (2022). A deep learning-based approach for inappropriate content detection and classification of youtube videos. *IEEE Access*, *10*, 16283-16298.
18. Dhivya, M. N., & Banupriya, M. S. (2020). Network security with cryptography and steganography. *International Journal of Engineering Research & Technology (IJERT)*, *8*(3), 1-4.
19. Do Xuan, C., Nguyen, H. D., & Tisenko, V. N. (2020). Malicious URL detection based on machine learning. *International Journal of Advanced Computer Science and Applications*, *11*(1).
20. Rodríguez, G. E., Torres, J. G., Flores, P., & Benavides, D. E. (2020). Cross-site scripting (XSS) attacks and mitigation: A survey. *Computer Networks*, *166*, 106960.
21. Alhassun, A. S., & Rassam, M. A. (2022). A combined text-based and metadata-based deep-learning framework for the detection of spam accounts on the social media platform twitter. *Processes*, *10*(3), 439.
22. Bhardwaj, A., Al-Turjman, F., Sapra, V., Kumar, M., & Stephan, T. (2021). Privacy-aware detection framework to mitigate new-age phishing attacks. *Computers & Electrical Engineering*, *96*, 107546.
23. Amritha, P. P., Sethumadhavan, M., Krishnan, R., & Pal, S. K. (2019). Anti-forensic approach to remove stego content from images and videos. *Journal of Cyber Security and Mobility*, 295-320.
24. Jiang, J. A., Scheuerman, M. K., Fiesler, C., & Brubaker, J. R. (2021). Understanding international perceptions of the severity of harmful content online. *PloS one*, *16*(8), e0256762.
25. Latchoumi, T. P., Reddy, M. S., & Balamurugan, K. (2020). Applied machine learning predictive analytics to SQL injection attack detection and prevention. *European Journal of Molecular & Clinical Medicine*, *7*(02), 2020.
26. https://www.kaggle.com/datasets/saurabhshahane/classification-of-malwares
27. https://www.kaggle.com/datasets/matthewfields/malware-as-images