2024 4th International Conference on Pervasive Computing and Social Networking (ICPCSN) | 979-8-3503-8634-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICPCSN62568.2024.00010

Improved Generative Adversarial Network for Phishing Attack Detection

Shammi L
Research Scholar, Department of CSE,
Presidency University, Bengaluru, Karnataka
shammiazhar@gmail.com

Dr. C Emilin Shyni
School of CSE and IS
Presidency University, Bengaluru, Karnataka
emilinshyni2020@gmail.com

***Abstract*— A general form of attack that is happening over the internet is called phishing which could lead to identity theft and financial damages. Due to the increase of online electronic services and payment systems, the demand for accurate phishing detection tools has risen in recent times. However, the models often lead to high false detection rates. This research work introduces an Improved Generative Adversarial Network-based phishing attack detection, which has mainly two stages such as pre-processing and attack detection. Initially, for data pre-processing, a min-max normalization process is used. Following that, the attack detection process is done via Improved GAN, where a new discriminator loss function is adopted to enhance the detection performance. Finally, the performance of the proposed work is validated in terms of different performance measures.**

Keywords—*Phishing attack, Improved Generative Adversarial Network (IGAN), Attack detection, Min-max normalization.*

I. INTRODUCTION

Generally, phishing is a kind of social engineering attack. The users are tricked into disclosing their credentials like credit cards and passwords, along with other confidential information. It shows a great impact on individuals and industries in terms of personal and financial data. The increasing count of phishing attacks leads to the requirement for efficient attack detection approaches. Consequently, the victims are warned, while become a phishing campaign's target, to avoid any sensitive data loss [1] [2] [3].

Phishing is a dangerous cyber-attack that can result in financial losses and identity theft. The proliferation of online electronic services and payment systems has increased the need for high-accuracy phishing detection software. Because most phishing detection methods rely on elements found in webpage content, it requires crawling the website and using third-party services. Relying solely on webpage content-related criteria results in high false detection rates and poor detection accuracy. Deep learning has gained popularity recently as a method for identifying fraudulent websites. This phishing attack detection approach includes two kinds of approaches, which are software-based and human-based approaches. To enhance the end-user's knowledge and assist them in making good decisions while facing a phishing website [4] [5], human-based models are utilized. On the other hand, diverse techniques are adopted by the software-based models to decide whether a website is legitimate or phishing, without the interference of the end-user. The latter model has 5 types, which were, heuristic, visual similarity, blacklist/whitelist, ML, and DL [6] [7] [8].

The blacklist/whitelist-dependent approaches are mainly based on a list of recognized phishing websites, which includes data such as IP addresses, phishing URLs, and so on. Constantly updating this list is also crucial in the blacklist/whitelist-dependent approaches [9] [10]. Using diverse features, a phishing webpage's visual similarity is compared with its respective legitimate webpage, in this visual similarity-based method [11][12]. A webpage has been considered phishing if the similarity is greater than the preset threshold. The heuristic-dependent methods mainly rely on the characteristics of a phishing web page, experts' prior knowledge, or the similarity among the phishing web pages. Phishing detection has been considered a binary classification problem in the case of ML approaches. In the case of DL approaches, the detection is conducted with three steps, designing of DL model, model input selection, and analyzing the features used to categorize the websites [13][14][15]. However, only limited work is been done by GAN [16] architecture. On the other hand, the generative adversarial network (GAN) has received less attention. To address this, this work concentrates on proposing an improved GAN for automated phishing attack detection.

- To propose an effective phishing attack detection system using Improved GAN architecture adopting with trimmed factor-based loss function for accurate detection outcomes.

This work has been arranged as follows: a few recent publications related to Phishing attack detection have been reviewed in section II, section III details the proposed IGAN-based phishing attack detection system, section IV provides this work's implementation outcomes, and this work concludes in section V.

II. LITERATURE SURVEY

Some recent publications related to Phishing attack detection have been reviewed below

A phishing detection model called PDGAN was developed by Saad Al-Ahmadi *et al* [17] in 2022. For attaining reliable performance, this PDGAN depends on a website's URL. For synthetic phishing URL generation, LSTM was utilized. To categorize whether that URL was legitimate or phishing, CNN was utilized. For experiments, DomCop and Phish Tank datasets are utilized and the detection accuracy and precision of 89.58% and 91.02%were attained by PDGAN.

979-8-3503-8634-9/24/$31.00 ©2024 IEEE
DOI 10.1109/ICPCSN62568.2024.00010

An ML-based phishing detection model was proposed by Ala Mughaid *et al* [18] in 2022. Initially, the data set was partitioned into training and testing data. With the training the data, the model was trained, and using the test data, the results were validated. With 3 diverse datasets, this approach was conducted, and based on the email text characteristics and other features, the attacks are categorized into non-phishing or phishing. With a boosted decision tree, accuracy of 0.88, 1.00, and 0.97 was attained.

In 2023, Muhammad Waqas Shaukat *et al* [19] proposed a multilayered ML-based approach. For phishing detection, this approach categorizes the websites and advertisements into phishing and legitimate websites. Based on the prominent features, diverse ML and DL models are trained. As per the URL features, the websites are categorized as phishing or legitimate by this predictive model. Outstanding results were attained by these models.

For fraudulent URL identification, SK Hasane Ahammad *et al* [20] proposed an ML-based model in 2022. In this work, LGBM, DT, SVM, and LR are utilized. The URL's domain-based and linguistic properties are investigated. From these algorithms, LGBM attained the best outcomes. However, for accuracy enhancement, more features needed to be added since some URLs in the dataset were not present in the database.

An ML-based phishing detection system was proposed in 2023, by Abdul Karim *et al* [21]. This model was named the LSD Ensemble model. This model was compared with DT, RL, SVM, GBM, SVM, K-Neighbour, NB, and hybrid (LR+SVM+DT) models. Higher performance was attained by this LSD Ensemble model. Furthermore, this model utilized the cross-fold validation along with canopy feature selection and Grid search hyperparameter optimization approach.

TABLE I: ADVANTAGES AND CHALLENGES OF SOME RECENT PHISHING ATTACK DETECTION MODELS

Citation	Techniques	Advantages	Challenges
Saad Al-Ahmadi et al [17]	GAN, LSTM,and CNN	Higher detection accuracy (89.58%) and precision (91.02%) were obtained	Depending on the batch size, the accuracy and loss ratings also change.
Ala Mughaid*et al* [18]	ML	For 50 features, it provides the highest accuracy	Limitation in finding the predefined dataset and while using 20 features, it cannot find the phishing attacks
Muhammad Waqas Shaukat *et al* [19]	ML	The promising results were attained by XGBoost	To provide a more robust detection process, this approach should be further optimized
SK Hasane Ahammad *et al* [20]	ML models (LGBM,DT, LR and SVM)	Best outcomes were provided by LGBM	For accuracy enhancement, more features needed to be added since some URLs in the dataset were not present in the whois database
Abdul Karim *et al* [21]	Hybrid LSD model	Effectively detecting the Phishing attacks	To prevent and detect the attacks more effectively, this should be merged with list - based ML-dependent systems.

A.Problem Statement

Recently, governments, internet users, and service-providing organizations have faced an important attack named phishing attack. The client's sensitive data is collected by the attackers with the utilization of fake websites or spoofed emails. Diverse researches are conducted related to the detection of phishing attacks; however, also face challenges. Recently, the GAN [17] model was utilized for the phishing attack detection. Although higher detection accuracy (89.58%) and precision (91.02%) were obtained by this approach, the accuracy rating was greatly dependent on the batch size. Furthermore, DL models were utilized in [18], which provides higher accuracy for more features than was while using 50 features. Although it can provide sufficient accuracy for 20 features, these outcomes are not effective enough for phis hing email detection. It's another drawback was finding the attack in the predefined dataset. For effective phishing attack detection, ML models were utilized in [19]. This work proved that promising results were attained by XGBoost. Although this model was optimized in terms of computational efficiency, to provide a more robust detection process, this approach should be further optimized. Similarly, ML models such as LGBM,DT, LR, and SVM are utilized in [20]. Although the best outcomes were provided by LGBM, for accuracy enhancement, more features needed to be added since some URLs in the dataset were not present in the Who is database. Moreover, the Hybrid LSD model [21] was developed to effectively detect Phishing attacks. However, to prevent and detect the attacks more effectively, this should be merged with list-based ML-dependent systems (Table I). To tackle these shortcomings, new methods for phishing attacks should be developed. Consequently, an Improved GAN-based phishing attack detection model is developed in this work, which is detailed below.

III. PROPOSED IGAN-BASED PHISHING ATTACK DETECTION

Nowadays, for social interactions, the internet has been considered a powerful channel. Since, people are highly dependent on digital platforms, which pave the way for data theft or attacks. From online platforms like online businesses, online classrooms, online banking, e-commerce, digital marketplaces, etc, the users' credentials are stolen by the attacker, and this is known as a phishing attack. For phishing webpage detection, many tools are introduced by researchers, which are white-list, antivirus software, and blacklist. To penetrate the cyber defense, attackers utilize creative ways. Some detection models require high computational power to compute the features obtained from different sources. Hence, there need for a robust knowledge-based model to categorize the attack and nonattack data. In this work, an IGAN-based phishing attack detection model is introduced, which includes two stages: pre-processing and attack detection. Initially, for pre-processing, a min-max normalization process is utilized.

Afterwards, an IGAN-based phishing attack detection process is done, which is detailed below and its architecture is displayed in Fig 1.

A. Pre-processing

Initially, the input data IP_D is subjected to the pre-processing phase, where min-max normalization is utilized. A general technique utilized in data pre-processing is named min-max normalization. This transforms a dataset's value into a common scale. To preserve the relationships among the original values, have selected the min-max normalization process In the work, min-max normalization [22] performed a linear transformation on the IP_D and obtained the scaled data in the range (0, 1), that is numerically displayed in Eq. (1).

$$IP_{DS} = \frac{A - A_{\min}}{A_{\max} - A_{\min}} \quad (1)$$

Here, the scaled data is denoted as IP_{DS} and A is the data. Furthermore, $A_{\max}$ and $A_{\min}$ denotes the maximum and minimum values of data. Subsequently, the obtained IP_{DS} is subjected to Improved GAN for detecting the attacks.

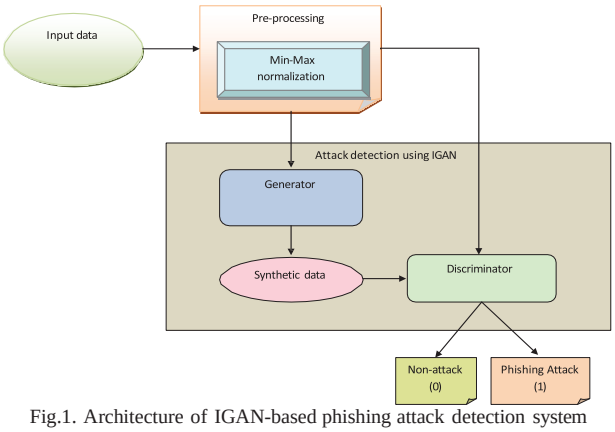


Fig.1. Architecture of IGAN-based phishing attack detection system

B. Attack detection using Improved GAN (IGAN)

The obtained IP_{DS} subjected to IGAN.IGAN is the unsupervised learning model that includes two NNs, a generator (GE) and a discriminator (DI). Both GE and DI models are probabilistic.

Generator (GE):

The IGAN's generator network includes the layer arrangement such as input layer, dense layer, leaky ReLU, reshape function, 2D convolution transpose, leaky ReLU, 2D convolution transpose, leaky ReLU, and afterwards a 2D convolution layer, that is displayed in Fig 2. Also, the reshape function has the size of (7, 7,128).

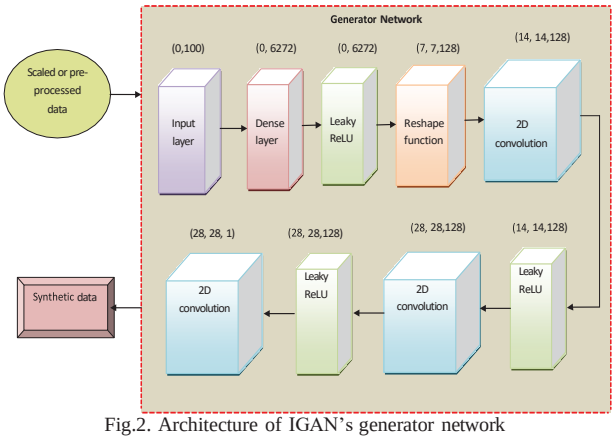


Fig.2. Architecture of IGAN's generator network

For capturing the data distribution, a generative model named generator is trained. From the learned distribution, the generator can output the generative and synthetic samples. To ensure the generated output sample's diversity, random noise(z) is present in this network. For learning the input data distribution, the joint probability distribution is applied by generative models using Eq. (2). Learning the data distribution facilitates the model to retrieve more prominent features and identify the data generation process. Therefore, artificial data having the identical distribution as the real data is provided by the generative models. These obtained synthetic data are plausible and dissimilar from the domain's real data.

$$p(IP_{DS}, Y) = p(IP_{DS}) \times p(Y) \quad (2)$$

Where, p is probability and IP_{DS} is the input and Y denotes the class labels.

Discriminator (DI): IGAN's discriminator network includes the following layer arrangement. Firstly, the input layer, 2D convolutional layer, LeakyReLU, 2D convolutional, LeakyReLU, 2D convolutional, LeakyReLU, flatten, dropout and dense layer. The activation utilized in the network is lambda. The discriminator network's architecture is seen in Fig 3.

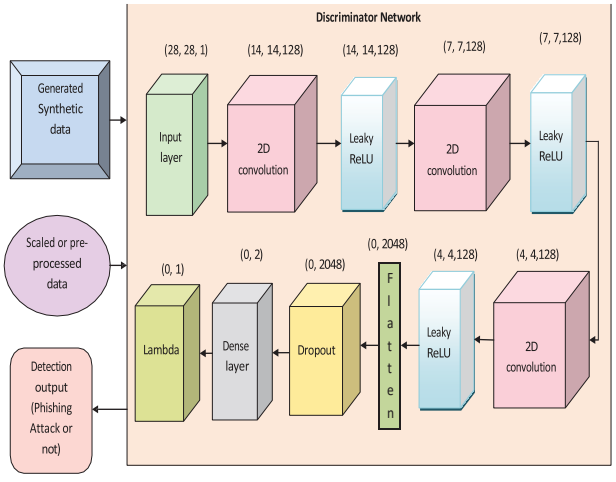


Fig.3. Architecture of IGAN's Discriminator network

A discriminative model has been trained to identify in which distribution the input data belongs. It helps to discover whether belong to real or artificial data distribution. Also, GAN's total performance can be enhanced by DI. GAN's training process has been considered as a min-max 2-player game.

To learn how to map input (IP_{DS})to their class labels (y), the conditional probability distribution is applied by the discriminative models using Eq. (3).

$$p(IP_{DS} \mid Y) = \frac{p(IP_{DS}, Y)}{p(Y)} \quad (3)$$

For maximizing the DI loss, the GE has been trained. The generator and discriminator losses of conventional GAN is in Eq. (4) and Eq. (5)

$$L^{(GE)} = \min\left[\log DI(IP_{DS}) + \log\left(1 - DI\left(GE(z)\right)\right)\right] \quad (4)$$

$$L^{(DI)} = \max\left[\log DI(IP_{DS}) + \log\left(1 - DI\left(GE(z)\right)\right)\right] \quad (5)$$

Here, $L^{(GE)}$ and $L^{(DI)}$ are generator and discriminator losses. Conventional discriminator loss function (in Eq. (5)), generally suffers from vanishing gradient issues and training instability. This will affect the detection performance of IGAN. For that reason, IGAN uses the trimmed factor-based loss function, which is mathematically expressed in Eq. (6). This trimming factor-based loss function (E_{TFL}) can provide accurate loss value.

$$E_{TFL} = -\frac{1}{H}\sum_{i=1}^{H} Q_{i:N} \quad (6)$$

where, H is the trimming factor that is considered during the training loss computation and it evaluates the smallest residual's quantity, N is the training set count.

This trimming factor-based loss function has additional functions based on constraints including data-specific characteristics and domain-specific knowledge. The model's ability to capture the relevant pattern in the data can be improved by this incorporation of a trimming factor-based loss function. Also, this loss function can help in ensuring the fair representation of all classes during training. For each observation, ordered losses are $Q_{1:N} \leq \ldots Q_{N:N}$, which can be written as in Eq. (7)

$$Q_i = \sum_{c=1} \lambda . PR_{ic} \log\left(\lambda T_{ic}\right) \quad (7)$$

Here, PR_{ic} symbolizes the predicted probabilities and C symbolizes the number of classes. Also, T_{ic} symbolizes the true values and λ denotes the robust factor, which can be evaluated using Eq. (8). Where δ =1.

$$\lambda = \delta\left(\left|PR_{ic} - T_{ic}\right| - \frac{1}{2}\delta\right) \quad (8)$$

This λ attains a smooth loss. Two versions of trimmed loss functions were tested in proposed work, which were when h=0.7 and h=0.9.

Finally, the discriminator provides the detection output based on the class labels, which might be 0 or 1 which means the class 0 indicates a non-attack and 1 indicates a phishing attack,

IV. RESULTS AND DISCUSSION

A. Simulation Procedure

The proposed phishing attack detection was simulated using PYTHON, specifically with "PYTHON 3.7." Further, the processor utilized was "11th Gen Intel (IR) Core (TM) i5-1135G7 @ 2.40GHz 2.42 GHz. Additionally, the system had "16.0 GB" of installed RAM. Moreover, the evaluation of phishing attack detection was carried out using the Phishing Dataset for Machine Learning [23].

B. Dataset Description

The dataset comprises 48 features extracted from a collection of 5000 phishing web pages and 5000 legitimate web pages. These web pages were acquired during two periods: from January to May 2015 and from May to June 2017. To enhance the extraction process, an advanced technique utilizing the browser automation framework, specifically Selenium WebDriver, was employed. This method proves to be more accurate and resilient when compared to the parsing approach reliant on regular expressions.

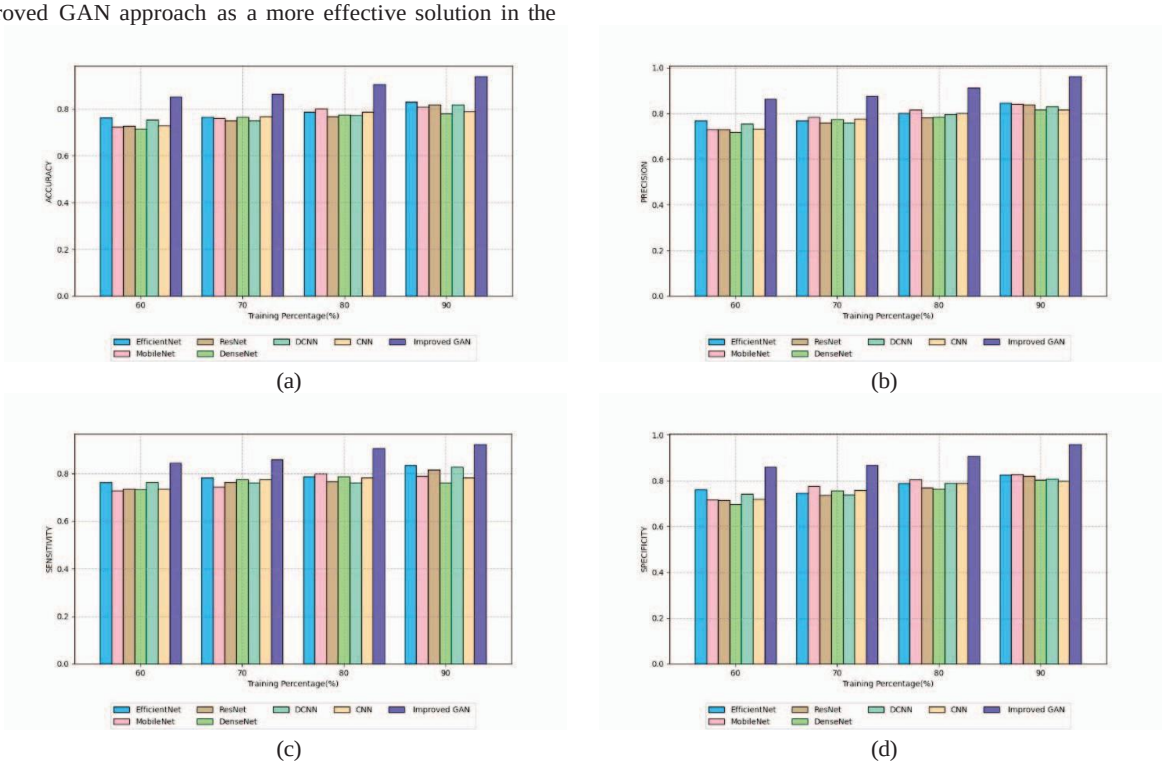
C. Performance Analysis

A thorough comparative analysis was conducted to evaluate the efficacy of the Improved GAN method in contrast to conventional approaches for detecting phishing attacks. This comprehensive assessment considered key metrics including Sensitivity, False Negative Rate (FNR), Negative Predictive Value (NPV), Specificity, F-measure, Precision, False Positive Rate (FPR), Matthews Correlation Coefficient (MCC), and Accuracy. Moreover, the performance of the Improved GAN scheme was compared with that of traditional approaches, including EfficientNet, MobileNet, ResNet, DenseNet, DCNN, and CNN [17].

D. Comparative Analysis on Positive Metric

In the context of phishing attack detection, a comparative analysis has been conducted to evaluate the Improved GAN strategy against conventional methods. Figure 4 illustrates the positive metric, a crucial factor in gauging the efficacy of the detection technique. The objective is to maximize the positive metric value for the efficient and effective identification of phishing attacks. To benchmark the Improved GAN strategy, a thorough comparison is made with established models such as EfficientNet, MobileNet, ResNet, DenseNet, DCNN, and CNN. Specifically, when considering a training percentage of 90, the performance of the Improved GAN model stands out with a notable detection accuracy of 0.938. In contrast, conventional methods demonstrate comparatively lower accuracy ratings: EfficientNet at 0.823, MobileNet at 0.809, ResNet at 0.816, DenseNet at 0.783, DCNN at 0.837, and CNN at 0.796, respectively. Furthermore, an additional noteworthy observation is the exceptional sensitivity achieved by the Improved GAN scheme, reaching a peak value of 0.895 at a training percentage of 80. This superior sensitivity

performance surpasses that of established models such as Efficient Net, Mobile Net, ResNet, DenseNet, DCNN, and CNN, respectively. The results underscore the potential of the Improved GAN approach as a more effective solution in the realm of phishing attack detection, showcasing its ability to outperform established models under the specified training conditions.



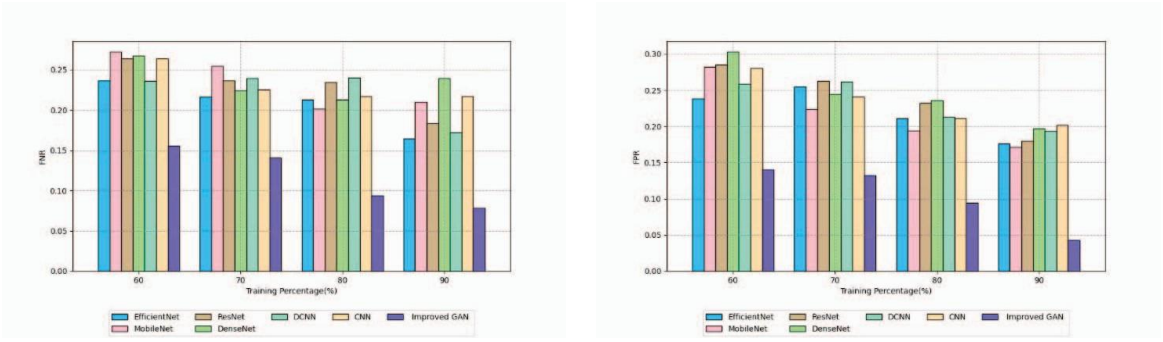
(a) (b)

(c) (d)

Fig.4. Assessment of Improved GAN and conventional strategies using Positive Metric

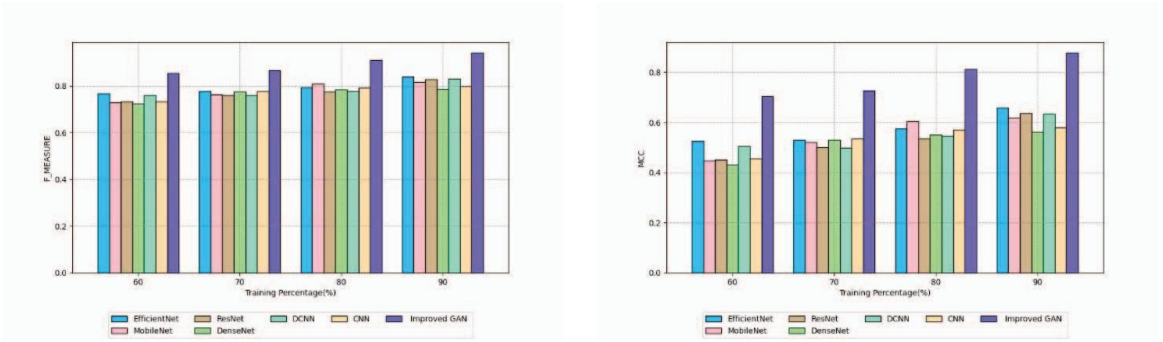
E. Comparative Analysis on Negative and Other Metrics

In the realm of phishing attack detection, the evaluation extends beyond positive metrics, encompassing negative metrics and other performance metrics to provide a comprehensive understanding of the Improved GAN model's efficacy. In this analysis, the Improved GAN model's performance is systematically contrasted with that of conventional methods, with a focus on minimizing negative values while concurrently enhancing other metrics. Figure 5 and Figure 6 serve as visual representations, highlighting the comparison of the Improved GAN model with the established methodologies. The objective is to showcase the Improved GAN model's proficiency in mitigating false negatives and optimizing other metrics critical to the accurate identification of phishing threats. Furthermore, the FPR of the Improved GAN scheme stands at 0.132 for a training percentage of 70. In contrast, established models exhibit higher FPR values, with Efficient Net at 0.254, Mobile Net at 0.223, ResNet at 0.262, Dense Net at 0.244, DCNN at 0.261, and CNN at 0.240, respectively. This discrepancy underscores the superior performance of the Improved GAN scheme in minimizing false positives compared to conventional methods. Similarly, the NPV of the Improved GAN approach is 0.913, whereas the Efficient Net, Mobile Net, ResNet, DenseNet, DCNN and CNN yielded the greatest NPV ratings of 0.813, 0.775, 0.795, 0.745, 0.803 and 0.762, respectively. Furthermore, the Improved GAN approach consistently achieved superior ratings compared to conventional methodologies across all metrics, ensuring precise detection of phishing attacks.

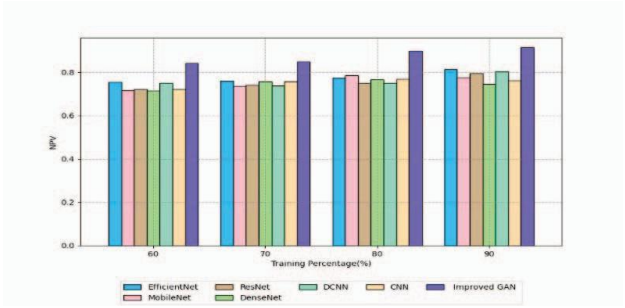


(a) (b)

Fig.5. Assessment on Improved GAN and conventional strategies using Negat ive Metric



(a) (b)



(c)

Fig.6. Assessment of Improved GAN and Conventional Strategies using Other Metric

F. Statistical Analysis on Accuracy

To ensure accurate results, each method undergoes a thorough statistical evaluation, including a comprehensive examination of key statistical parameters such as "Maximum, Minimum, Mean, Standard Deviation, and Median." Table II provides a comprehensive statistical assessment comparing the Improved GAN model with Efficient Net, Mobile Net, Res Net, Dense Net, DCNN, and CNN for phishing attack detection. Notably, in terms of the maximum statistical metric, the Improved GAN scheme demonstrates a remarkable accuracy of 0.938, meanwhile, the Efficient Net is 0.830, Mobile Net is 0.808, Res Net is 0.818, Dense Net is 0.780, DCNN is 0.818 and CNN is 0.790, respectively. Likewise, across most statistical metrics, the Improved GAN model consistently achieved higher accuracy ratings.

TABLE II: STATISTICAL EVALUATION OF ACCURACY

Statistical Metrics	Efficient Net	Mobile Net	ResNet	Dense Net	DCNN	CNN	Improved GAN
Standard Deviation	0.027	0.034	0.034	0.026	0.027	0.025	0.034
Mean	0.786	0.773	0.765	0.759	0.774	0.768	0.890
Minimum	0.763	0.723	0.726	0.716	0.750	0.728	0.852
Median	0.777	0.781	0.759	0.771	0.763	0.777	0.885
Maximum	0.830	0.808	0.818	0.780	0.818	0.790	0.938

V. CONCLUSION

A social engineering attack, which targets users' emails to steal sensitive and confidential information, is known as a phishing attack. It can be utilized as an enormous attack's part that is launched to invade government or corporate networks. Numerous anti-phishing attacks have been developed by researchers, but are inaccurate and inefficient. Consequently, an IGAN-based phishing attack detection system was introduced in this work. Initially, for data pre-processing, min-max normalization process was utilized. Afterwards, a phishing attack detection process takes place, which makes use of IGAN to provide accurate and efficient attack detection through the proposed trimming factor-based loss function. In the future, the suggested method's next steps will require classifying into two classes in order to improve the outcomes.

REFERENCES

[1] D. Rathee, and S. Mann 2022, Detection of E-mail phishing attacks–using machine learning and deep learning. *International Journal of Computer Applications*, vol. 183, no 1, pp.1-7.

[2] E. Benavides-Astudillo, W. Fuertes, S. Sanchez-Gordon, D. Nuñez-Agurto, and G. Rodríguez-Galán 2023. A Phishing-Attack-Detection Model Using Natural Language *Processing and Deep Learning. Applied Sciences*, vol. 13, no. 9, pp.1-23.

[3] U.A. Butt, R. Amin, H. Aldabbas, S. Mohan, Alouffi, and A. Ahmadian, 2023. *Cloud-based email phishing attack using machine and deep learning algorithms. Complex & Intelligent Systems*, vol. 9, no. 3, pp.3043-3070.

[4] M.J. Nabet, and L. E. George, 2023. Phishing Attacks Detection by Using Support Vector Machine. Journal of Al-Qadisiyah for Computer Science and Mathematics, vol. 15, no. 2, pp.180 .

[5] G. Mohamed, J. Visumathi, M. Mahdal, J. Anand, and M. Elangovan 2022, An effective and secure mechanism for phishing attacks using a machine learning approach. *Processes*, vol. 10, no. 7, pp.1-14.

[6] S. Hossain, D. Sarma, and R. J. Chakma 2020, Machine learning-based phishing attack detection. International Journal of Advanced Computer Science and Applications, vol. 11, no. 9.pp.378-388.

[7] Y. Wang, W. Ma, H. Xu, Y. Liu, and P. Yin, 2023. A Lightweight Multi-View Learning Approach for Phishing Attack Detection Using Transformer with Mixture of Experts. *Applied Sciences*, vol. 13, no. 13, pp.1-17.

[8] T. Choudhary, S. Mhapankar, R. Bhddha, A. Kharuk, and R. Patil, 2023. A Machine Learning Approach for Phishing Attack Detection. *Journal of Artificial Intelligence and Technology*, vol.3, no. 3, pp. 108–113.

[9] Y.A. Alsariera, A.O. Balogun, V. E. O. H. Adeyemo, Tarawneh, and H.A. Mojeed, 2022. Intelligent tree-based ensemble approaches for phishing website detection. *J. Eng. Sci. Technol*, 17, pp.563-582.

[10] M. Sánchez-Paniagua, E.F. Fernández, E. Alegre, W. Al-Nabki, and V. Gonzalez-Castro 2022, Phishing URL detection: A real-case scenario through login URLs. *IEEE Access*, vol.10, pp.42949-42960.

[11] A.T.G. Tapeh, and M. Z. Naser 2023, Artificial intelligence, machine learning, and deep learning in structural engineering: a scientometrics review of trends and best practices. Archives of Computational Methods in Engineering, vol. 30, no. 1, pp.115-159.

[12] F. Salahdine, Z. El Mrabet, and N. Kaabouch, 2021, December. Phishing Attacks Detection A Machine Learning-Based Approach. *In 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) (pp. 0250-0255). IEEE.*

[13] A. A. Orunsolu, A. S. Sodiya, and A.T. Akinwale, A predictive model for phishing detection 2022. *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no.2, pp.232-247.

[14] M..Bhurtel, Y.R. Siwakoti, and D. B. Rawat, Phishing Attack Detection with ML-Based Siamese Empowered ORB Logo Recognition and IP Mapper. *In IEEE INFOCOM 2022-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (pp. 1-6). IEEE. 2022.

[15] G. Sonowal, and K.S. Kuppusamy 2020, PhiDMA–A phishing detection model with multi-filter approach. Journal of King Saud University-Computer and Information Sciences, vol. 32, no. 1, pp.99-112.

[16] N. Elaraby, S. Barakat, and A. Rezk 2022, " A conditional GAN-based approach for enhancing transfer learning performance in few-shot HCR tasks". *Scientific Reports*, vol. 12, no.1, p.16271.

[17] S. Al-Ahmadi, A. Alotaibi, and O. Alsaleh 2022, PDGAN: Phishing detection with generative adversarial networks. *IEEE Access*, vol. 10, pp.42459-42468.

[18] A. Mughaid, S. AlZu'bi, A. Hnaif, S. Taamneh, A. Alnajjar, and E.A. Elsoud 2022, An intelligent cyber security phishing detection system using deep learning techniques. *Cluster Computing*, vol. 25, no.6, pp.3819-3828.

[19] M.W. Shaukat, R. Amin, M.M.A. Muslam, A.H. Alshehri, and J. Xie, 2023. A hybrid approach for alluring ads phishing attack detection using machine learning. *Sensors*, vol. 23, no. 19, pp.1-27.

[20] S.H. Ahammad, S.D. Kale, G.D. Upadhye, S.D. Pande, E.V. Babu, A.V. Dhumane, and M.D.K.J. Bahadur2022, Phishing URL detection using machine learning methods. *Advances in Engineering Software*, vol. 173, no.103288.

[21] A. Karim, M. Shahroz, K. Mustofa, S.B. Belhaouari, and S.R.K. Joga, 2023. Phishing Detection System Through Hybrid Machine Learning Based on URL. *IEEE Access*, 11, pp.36805-36822.

[22] https://www.geeksforgeeks.org/data-normalization-in-data-mining/

[23] https://www.kaggle.com/datasets/shashwatwork/phishing-dataset -for-machine-learning?resource=download