

Artificial intelligence-powered intelligent reflecting surface systems countering adversarial attacks in machine learning

Rajendiran Muthusamy¹, Charulatha Kannan², Jayarathna Mani¹, Rathinasabapathi Govindharajan³,
Karthikeyan Ayyasamy¹

¹Department of Computer Science and Engineering, Faculty of Computer Science and Engineering, Panimalar Engineering College, Chennai, India

²Department of Artificial Intelligence and Data Science, Panimalar Engineering College, Chennai, India

³Department of Mechanical Engineering, Panimalar Engineering College, Chennai, India

Article Info

Article history:

Received May 6, 2023

Revised Sep 16, 2023

Accepted Sep 27, 2023

Keywords:

6G

Adversarial machine learning

Intelligent reflecting surfaces

Neural network

Next generation networks

ABSTRACT

With the increase in the computation power of devices wireless communication has started adopting machine learning (ML) techniques. Intelligent reflecting surface (IRS) is a programmable device that can be used to control electromagnetic wave propagation by changing the electric and magnetic values of its surface. State-of-the-art ML especially on deep learning (DL)-based IRS-enhanced communication is an emerging topic. Yet while integrating IRS with other emerging technologies possibilities of adversarial data creaping is high. Threats to security, their mitigation, and complexes for AI-powered applications in next generation networks are continuously emerging. In this work the ability of an IRS enhanced wireless network in future-generation networks to prevent adversarial machine-learning attacks is studied. The artificial intelligence (AI) model is used to minimize the susceptibility of attacks using defense distillation mitigation technique. The outcome shows that the defensive distillation technique (DDT) increases the strength and performance by around 22% of the AI method under an adversarial attack.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Rajendiran Muthusamy

Department of Computer Science and Engineering, Faculty of Computer Science and Engineering

Panimalar Engineering College

Poonamallee, Chennai, Tamil Nadu, India

Email: mrjendiran@panimalar.ac.in

1. INTRODUCTION

Next generation networks called NextG or 5G and 6G, are gaining more attention in both industry and academia. Consumers are expecting a high demand and new ways of communication. Based as a study by the international telecommunication union, mobile network traffic on 5th or 6th generation future networks will constantly increase year over year using thousands of pentabytes [1], [2]. The principle NextG networks is to transmit data immediately with least amount of delay between hardware and software devices and is commonly used in fields such as e-health medical services, cloning, artificial authenticity, various autonomous vehicles and online e-learning [3]. Next generation technologies are also used to enhance computing and communications systems. Artificial intelligence (AI) is one of the strong platforms that is very important in developing inventory models in the next generation network [4], [5].

An intelligent reflecting system (IRS) upgraded with multiple input and multiple output (MIMO) uses millimeter waves and is a powerful and efficient method in terms of channel capacity and data transmission ratio. It is also capable of reconfiguring wireless systems to obtain more concentration. IRS

utilizes a huge amount of minimum-cost passive send-back elements whose signals constructively add to the destination network, improving the output of the wireless communication networks. The AI model reduces its effective training, despite the various tools such as cyber security and AI, yet metamorphic and polymorphic security attacks. These adversarial attacks manipulate the AI model by intentionally mixing the original data with unwanted signals to the dataset and misleading it [6].

In this article, an AI-IRS system is proposed for next generation networks to reduce the vulnerability to a minimum level in the academic and business environment [7], [8]. This involves: i) calculating the susceptibilities of the AI methods of the IRS system by the adversarial attacks using fast gradient sign method (FGSM) and basic iterative method (BIM); ii) proposing a defensive distillation mitigation algorithm to improve the robustness and efficiency of the AI-model on the IRS system; and iii) training the AI-IRS systems to produce and maintain robust output data under undefended and defended methods using FGSM and BIM adversarial attacks.

2. METHOD

2.1. Intelligent reflecting surface wireless communication

Wireless communication quality can be enhanced using IRS wireless communication system which significantly improves the efficiency of communication between a sender and receiver. The destination receives both the line of sight (LOS) waves from the LOS connection and constructive send-back signals from the IRS recipient during idle time [9]. IRS can improve communication systems by dynamically changing wireless channels and adjusting the signal reflection surfaces via a large number of inexpensive passive reflecting devices. Though IRS-supported hybrid wireless network with passive and active components promises to achieve long-term and cost-effective capacity growth, it needs to overcome certain obstacles such as channel estimation, deployment, and reflection optimization [10].

This suggests that machine learning (ML) model has to be trained to detect the domain signifiers to expect the possible rate with each IRS interaction communication route. This can be achieved by the current developments in deep learning in which, the transmitter should reflect the sent data to the receiver and the IRS interaction route should be compatible with the highest expected realistic rate to be used. The method is referred to as the AI-method on the IRS system, this work where its weakness is examined and evaluated using defensive distillation mitigation strategy [11], [12].

2.2. Adversarial machine learning

Adversarial ML is used in a variety of applications and is primarily used to implement malicious attacks or reasons for ML model malfunctioning [13]. The principle is to train the models to automatically understand the original designs of the working procedure and relationships in data using the trained algorithms [1]. Post training is mostly used to calculate and analyze the outlines in given information [14]. Figure 1 shows the steps involved in wrong prediction due to attack on machine learning technique. The precision range of the trained model is important for obtaining a better outcome, which is addressed as a generalization. The various types of adversarial machine-learning attacks include data evasion, poisoning and model attacks [15]. Adversarial ML methods are used to finalize, locate adversaries and produce planned betrayals of the ML model.

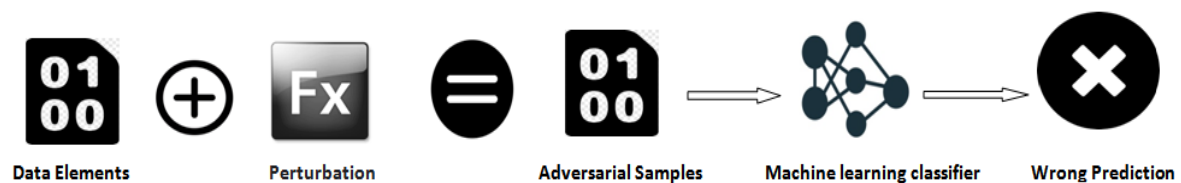


Figure 1. Adversarial attacks on data

The sample model input should confuse the model by executing an invalid classification with the given data that can be used to operate certain blind dots in image classifiers [16], [17]. This article's goal is to examine the most recent adversarial ML techniques to create and identify adversarial samples. Both targeted and non-targeted evasion attempts aim to persuade models to incorrectly identify malicious examples as valid data points. Targeted attacks attempt to persuade ML models to include adversaries in a special target model. Non-targeted attacks are designed to force ML models to order the adversarial example as a different model than reality [18]. The goal of data poisoning is to generate false data points that will be used to train ML

models in producing the desired results. Data poisoning can be used to produce the desired results using ML methods. Some examples of adversarial attacks are FGSM and BIM.

2.2.1. Fast gradient sign method

FGSM is a one-step attack in which the perturbation is added in a single step rather than over a loop. The fast gradient sign method involves the following three steps: first step is to compute the loss function and forward propagation and next step involves the calculation of gradient based on the pixels of the image and finally forward the image pixels a little bit in the direction of the estimated gradients to increase the loss in the previous steps [19].

A negative likelihood loss technique is applied to determine how closely the model's prediction matches the actual class. The computation of the gradients concerning the image pixels is unusual. Gradients are used in neural network training to determine the direction in which weights need to be changed to reduce the loss values. As an alternative, in this case, input image pixels are moved in the gradient's direction to increase the loss value. Back propagating the gradients from the start to the weight is the most commonly used method when training neural network to determine the direction by which a specific weight is altered deep in the neural network. In such situations, a similar idea [20] the gradients being returned to the input image from the output layer is applied.

The following mathematical formula is given to move the weights to reduce the loss value in neural network training:

$$updated_weights = previous_weights - learning_ratio \times Xgradients \quad (1)$$

the following mathematical formula is used to increase the loss and move the pixel values of the image:

$$new_pixels = old_pixels + epsilon * gradients \quad (2)$$

furthermore, the following algorithm is applied for perturbation in the fast gradient sign method.

$$X^{adv} = X + \epsilon \cdot sign(\nabla_x k(x, y_{true})) \quad (3)$$

Where X^{adv} is the adversarial image, ϵ is the perturbation and $(\nabla_x k(x, y_{true}))$ is the first derivative of the loss function concerning the input x . In the case of deep neural networks, this can be calculated using the back-propagation technique.

The following equation is used for targeted FGSM attacks:

$$X^{adv} = X - \epsilon \cdot sign(\nabla_x k(x, y_{target})) \quad (4)$$

X^{adv} is equal to the negative of k . In this case of targeted attacks, the loss function between the targeted class and the predicted class is minimized, whereas an untargeted attack maximizes the loss function between the predicted class and the true class [21].

2.2.2. Basic iterative method

ML algorithms iteratively study the data that permits the machine to find the hidden forms within the data [22]. The objective of a basic iterative algorithm is to find the best solution from the data set. These algorithms learn from previous experience that consistent and repeatable decisions are made to obtain the best solution [23].

The method can be repeated several times with small step sizes. This technique involves clipping the pixel values between the results in each phase to ensure that they are in the vicinity of the original image. That is within a certain range of the previous image's pixel value.

The following mathematical calculation is used for generating the perturbed pictures using this basic iterative method:

$$x_0^{adv} = x, x_{N+1}^{adv} = Clip_{x, \epsilon} \left\{ x_N^{adv} + \alpha \cdot sign(\nabla_x k(x_N^{adv}, y_{true})) \right\} \quad (5)$$

X^{adv} and x are the adversarial images at the i th step and input image respectively, k represents the means loss function, y_{true} is the output for input x , ϵ is the tuneable value and alpha is the step size. An overview of iterative algorithm is provided in Figure 2.

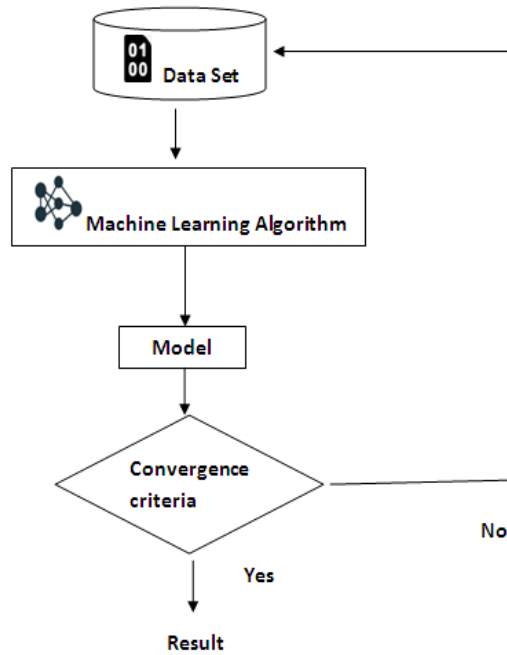


Figure 2. Overview of the iterative algorithm

3. WORK CONCEPT

In neural network architecture and defensive distillation technique (DDT), the input data received from the user devices is used to IRS prediction method. Defensive distillation training networks is covered using a defended model which has deep neural networks with large network and shallow neural networks with small neural network [24], [25]. The overall system design for the proposed AI-powered intelligent reflecting surface system is shown in Figure 3. The figure shows that during the prediction model training, a shallow neural network model, protected against adversarial ML attacks in mobile base stations. Adversarial attacks are applied in defended and undefended method to evaluate the methods under any attacks.

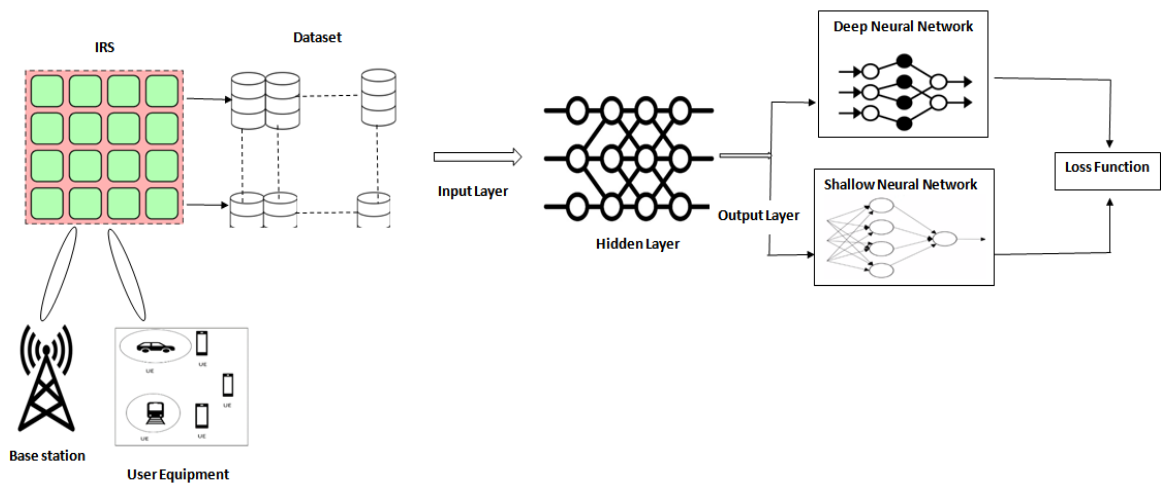


Figure 3. AI-powered intelligent reflecting surface systems

3.1. Neural network architecture

Neural network technique also called deep learning, the principles of human brain while processing data using a computer [26]. As shown in Figure 4, it uses interconnected nodes or neurons in a layered structure that resembles the human brain. The neural network input is a signal from the transmitter and

receiver of the uplink pilot. The neural network's output is a prediction score based on the input signals from the transmitter and receiver. Neural network consists of multiple layers of networks [27]. The output falls under multilayer layer perception in which inputs are processed by multiple layers of neurons.

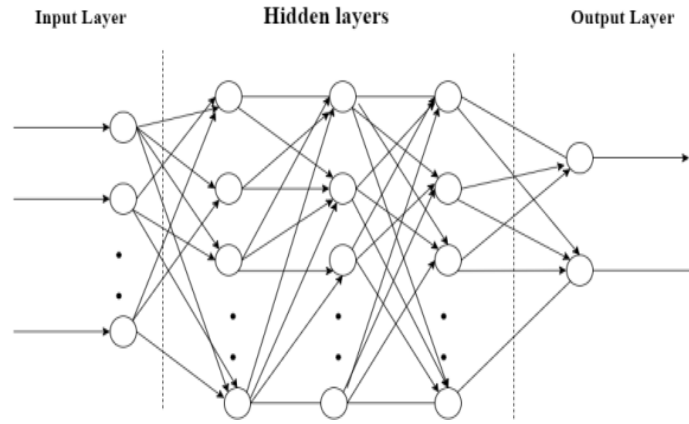


Figure 4. Neural network architecture

3.2. Defensive distillation technique

Defensive distillation technique is one of the most popular adversarial training method that adds flexibility to the classification process of an algorithm, making them less prone to attacks. DDT employs defensive knowledge distillation to train the model to be more powerful. Knowledge distillation was previously introduced by Catak *et al.* [28]. In this technique the knowledge of the master (densely connected neural network) is transferred to a slave (sparsely connected neural network). In knowledge distillation, the slave should perform similarly to the master by imitating the master's output, which causes soft labels to be used to train the slave network using the master node.

The workflow of the DDT consists of three steps. Step 1 is training the master model with the loss function for the classification of inputs. Step 2 again trains the previously trained master model with the defensive distillation method that produces a soft label and a cross-entropy loss function to generate the corresponding soft labels as outputs and the last step involves training the slave model using the soft labels from the previous process labels to produce a better, more robust and more accurate method. Algorithm 1 provides the defensive distillation technique used in this study to counter adversarial attacks in machine learning. The defensive distillation parameters of this study are provided Table 1. The loss function is defined as (6):

Algorithm 1. Defensive distillation

```

Function defensive distillation
    Call defensive ()
End Function
def defensive()
1: Read Dataset DS, Base Model  $M_T$ , Cross Entropy  $\lambda$ , Adversarial perturbation  $\epsilon$ 
2: Number of iterations N
3: Minimize the cross entropy loss  $L_{CE}$  on Dataset DS
4: Initialize the defensive distillation model  $M_{DS} = M_T, i = 0$ 
5: while  $i < N$  do
    Read the samples  $x$  and Labels  $y$ 
    Compute the following:
    Cross-entropy=  $L_{CE}(\theta)$ 
    Kullback Leibler Divergence=  $L_{KLD}(P_T(y/\theta), P_T(y))$ 
    Compute defend distillation loss:
     $L_{DF}(\theta) = (1 - \lambda)L_{CE}(\theta) + L_{KLD}(P_T(y/\theta), P_T(y))$ 
    Calculate FGSM and BIM with
    FGSM  $X_{adv} = X + \epsilon \times \text{sign}(\nabla_X l)$ 
    BIM  $X_{adv} = X_{adv} + \epsilon \times \text{sign}(\nabla_X l)$ 
    Update  $M_{DS}$ 
     $i \leftarrow i + 1$ 
6: Endwhile
7: return  $M_{DS}$ 

```

Table 1. Defensive distillation parameters

Parameter	Description
L_{DF}	Distillation loss function
L_{CE}	Cross-entropy loss
$P_T(y)$	Output of the shallow neural network model
$P_T(y/\theta)$	Output of the deep neural network model
L_{KL}	Kullback leiblerdivergence (KL) loss
λ	Parameter between KL divergence and cross entropy

$$L_{DF}(\theta) = (1 - \lambda)L_{CE}(\theta) + L_{KL}(P_T(y/\theta), P_T(y)) \quad (6)$$

4. EXPERIMENTAL AND RESULTS

AI-powered IRS methods evaluated using mean square error (MSE) algorithm. MSE scores are used to evaluate the model vulnerabilities under protected and unprotected conditions. The MSE is calculated as:

$$MSE = \frac{\sum(Y_t - \hat{Y}_t)^2}{n} \quad (7)$$

where: n denotes total number of samples, Y_t the actual data value and \hat{Y}_t the predicted data value.

The output represented in the form of bar plots (Figures 5 and 6) and histogram (Figures 7 and 8), which shows the MSE values for each adversarial ML attack on the protected and the unprotected systems. Table 2 shows that the prediction of performance outputs for the protected and unprotected AI-powered IRS method countering the attacks. The publicly available ray trace MIMO datasets are adopted to generate the training data and compare with the AI-powered IRS method. Based on the ray-tracing data obtained from the value ray-tracing simulation outline, the MIMO dataset parameter was used to build the MIMO channels.

The adversarial attack on the AI-powered method has become more popular with several attacks. BIM and FGSM types are used in this study to generate adversarial examples. The performance of each model was estimated through the MSE parametric.

The trained AI-powered IRS method was simulated using a python, tensor-flow framework executed using a Google Colab Tesla GPU with 16 GB memory. The adversarial input data were generated using the Cleverhans library. Figure 5 shows that MSE values for the selected attack method under the attack powers from 0.01 to 0.10. The MSE values are similar to both BIM and FGSM algorithms and is around 0.08 for all attack powers. Furthermore, MSE values for BIM attacks rise with increasing attack power, ranging between 0.008 and 0.009. The output shows that AI-powered models are considered vulnerable to adversarial attacks. Mitigation technique are broadly used to improve the robustness of AI-powered model against adversarial attack [29]. Based on this observation, the DDT was applied in this method to reduce the vulnerability against adversarial attacks. The performance of the AI-powered is estimated in terms of MSE after applying the mitigation method.

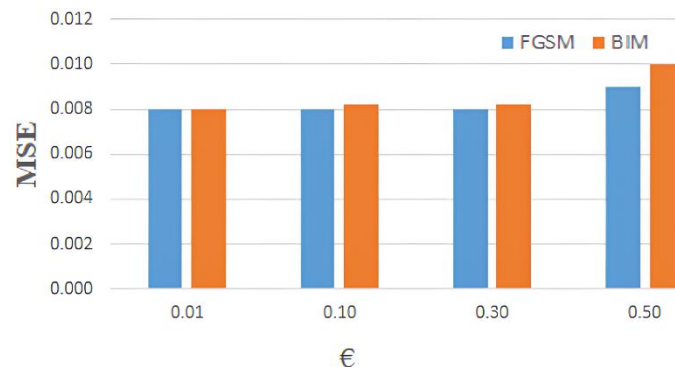


Figure 5. MSE value vs attack power (undefended models)

The MSE values against adversarial attacks range from 0.01 to 0.04 in Figure 6. The above cure depicts that the AI-powered is still prone to adversarial attacks, its robustness is better against adversarial

attacks. It was observed that the model can resist any attack under low attack power that is less than 0.30. Increasing the mean square value implies that high power attack is expected. The effect of the mitigation technique on the performance is not the same for all attacks. The MSE values can go between 0.001 and 0.003 under the FGSM and BIM attack respectively whereas under high attack power it goes up to 0.003 for BIM. On the other hand, the attack power under the FGSM attack is low when the mitigation technique is applied to the model. The output indicates that the defensive distillation model significantly contributes to the model's robustness against adversarial attacks.

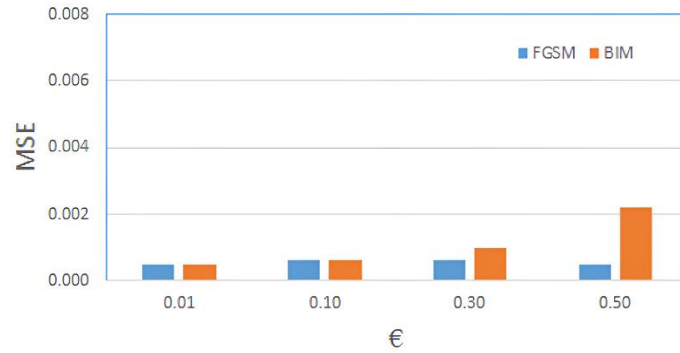


Figure 6. MSE value vs attack power (defended models)

Figure 7 examines the variation of MSE values for undefended under adversarial attacks. Based on the output data, the undefended AI models under FGSM adversarial attacks. Based on the output data, the undefended model corresponds to a moderate right skewed distribution which has a maximum out to the left of the distribution. The MSE values differ from 0.004 to 0.024 for all types of attacks. The percentage of high MSE values is lower than that of the undefended model. This indicates that the mitigation technique can significantly improve the method robustness under FGSM attacks.

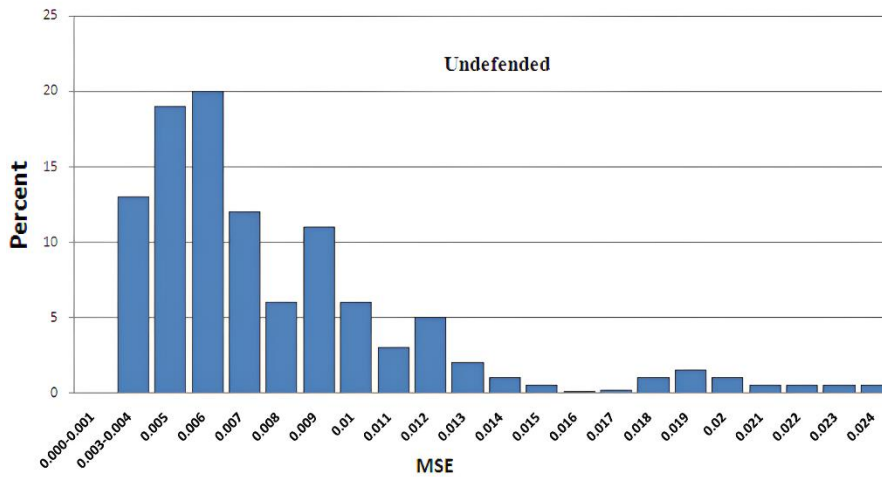


Figure 7. MSE values vs percentage (undefended model)

Figure 8 examines the distribution of MSE values for defended method under adversarial attacks. Based on the output data, the defended model to represent a slight right-skewed distribution such as the undefended model. Based on that it can be stated that the AI-powered model can accurately predict the target values. Against FGSM attacks defended holds are found to be more effective. This indicates that the robustness of the model can be dynamically enhanced with mitigation techniques against FGSM attacks.

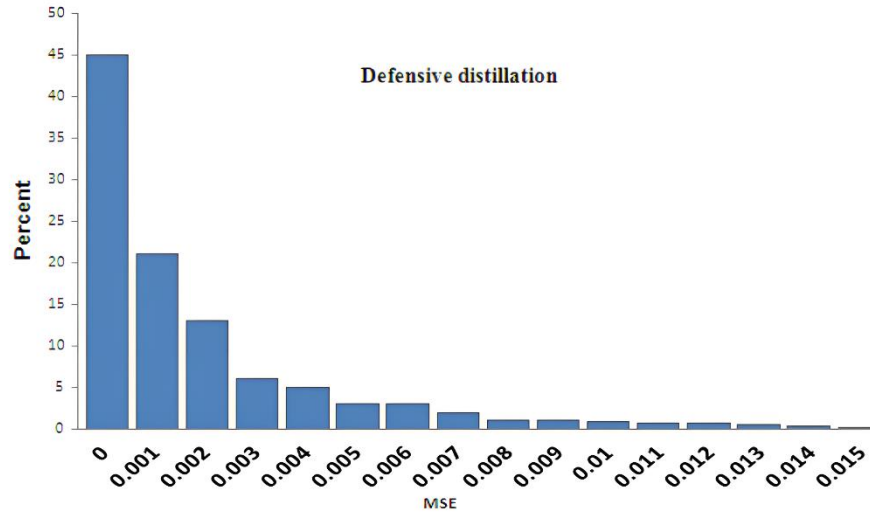


Figure 8. MSE values vs percentage (defensive distribution)

Table 2. Parameter setting

Parameters	Values
Frequency band (f)	28 GHz
Active base stations (bs)	4
Number of antennas (M)	{{(1, 16, 32); (1, 32, 64)}}
Receivers (rx)	R1100 to R1500
Transmitter (tx)	Row 900
	Column 95
Bandwidth(bw)	100 MHz
Number of subcarriers (sc)	512
OFDM sampling factor (sf)	1
OFDM limit (limit)	64
Number of channel paths (path)	1
Antenna spacing (λ)	0.5 λ

5. CONCLUSION

AI is the most important technology to the improvement of the performance of next generation network. This article examines the vulnerability of AI-powered IRS models against FGSM and BIM adversarial attacks. The impacts of the mitigation method such as defensive distillation improves the robustness in next generation networks. The output indicate that the AI-powered next generation networks are vulnerable to adversarial attacks. The overall result shows that BIM is the most effective adversarial attack (30%) on defended than undefended methods. The proposed defensive distillation mitigation method provides better results for defended FGSM attacks (22%) than undefended FGSM attacks. Future works can focus on vulnerabilities for various adversarial attacks such as Carlini and Wagner, momentum iterative method (MIM) and projected gradient descent (PGD) as well as the defensive distillation mitigation method.





REFERENCES

- [1] M. Zolotukhin, P. Miraghaei, D. Zhang, and T. Hamalainen, "On assessing vulnerabilities of the 5G networks to adversarial examples," *IEEE Access*, vol. 10, pp. 126285–126303, 2022, doi: 10.1109/ACCESS.2022.3225921.
- [2] I. Chatzigeorgiou, "The impact of 5G channel models on the performance of intelligent reflecting surfaces and decode-and-forward relaying," in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, Aug. 2020, pp. 1–4, doi: 10.1109/PIMRC48278.2020.9217321.
- [3] M. H. Shahriar, N. I. Haque, M. A. Rahman, and M. Alonso, "G-IDS: generative adversarial networks assisted intrusion detection system," in *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, Jul. 2020, pp. 376–385, doi: 10.1109/COMPSAC48688.2020.0-218.
- [4] A. Guesmi, K. N. Khasawneh, N. Abu-Ghazaleh, and I. Alouani, "ROOM: adversarial machine learning attacks under real-time constraints," in *2022 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2022, pp. 1–10, doi: 10.1109/IJCNN55064.2022.9892437.
- [5] S. Cho, T. J. Jun, B. Oh, and D. Kim, "DAPAS: denoising autoencoder to prevent adversarial attack in semantic segmentation," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–8, doi: 10.1109/IJCNN48605.2020.9207291.
- [6] X. Wang, J. Li, X. Kuang, Y. Tan, and J. Li, "The security of machine learning in an adversarial setting: a survey," *Journal of Parallel and Distributed Computing*, vol. 130, pp. 12–23, Aug. 2019, doi: 10.1016/j.jpdc.2019.03.003.




- [7] F. O. Catak, M. Kuzlu, E. Catak, U. Cali, and O. Guler, "Defensive distillation-based adversarial attack mitigation method for channel estimation using deep learning models in next-generation wireless networks," *IEEE Access*, vol. 10, pp. 98191–98203, 2022, doi: 10.1109/ACCESS.2022.3206385.
- [8] M. Swetha and M. Rajendiran, "Effective early stage detection of COVID-19 using deep learning," *Advances in Parallel Computing*, pp. 204–208, 2021.
- [9] B. Yang, X. Cao, C. Huang, C. Yuen, L. Qian, and M. D. Renzo, "Intelligent spectrum learning for wireless networks with reconfigurable intelligent surfaces," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 4, pp. 3920–3925, Apr. 2021, doi: 10.1109/TVT.2021.3064042.
- [10] J. Yu, X. Liu, Y. Gao, C. Zhang, and W. Zhang, "Deep learning for channel tracking in IRS-assisted UAV communication systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 7711–7722, Sep. 2022, doi: 10.1109/TWC.2022.3160517.
- [11] Q. Pan, J. Wu, X. Zheng, J. Li, S. Li, and A. V. Vasilakos, "Leveraging AI and intelligent reflecting surface for energy-efficient communication in 6g iot," *arXiv preprint arXiv:2012.14716*, 2020.
- [12] C. Sadu and P. K. Das, "A defense method against facial adversarial attacks," in *TENCON 2021 - 2021 IEEE Region 10 Conference (TENCON)*, Dec. 2021, pp. 459–463, doi: 10.1109/TENCON54134.2021.9707433.
- [13] A. Yadav, A. Upadhyay, and S. Sharanya, "An integrated auto encoder-block switching defense approach to prevent adversarial attacks," *arXiv preprint arXiv:2203.10930*, 2022.
- [14] K. P. Ashvitha, M. ShilpaArathi, M. R. Thamizhkanal, M. Rajendiran, and S. Malathi, "Hybrid segmentation and 3D modeling of pleural effusion on CT," *Proceedings of the International Conference for Phoenix on Emerging Current Trends in Engineering and Management (PECTEAM 2018)*, 2018, doi: 10.2991/pecteam-18.2018.29.
- [15] N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z. B. Celik, and A. Swami, "The limitations of deep learning in adversarial settings," in *2016 IEEE European Symposium on Security and Privacy (EuroS&P)*, Mar. 2016, pp. 372–387, doi: 10.1109/EuroSP.2016.36.
- [16] D. Madhubala, M. Rajendiran, and D. Elangovan, "A study on effective analysis of machine learning algorithm towards the women's safety in social media," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Nov. 2020, pp. 1151–1156, doi: 10.1109/ICECA49313.2020.9297386.
- [17] W. Xu, D. Evans, and Y. Qi, "Feature squeezing: detecting adversarial examples in deep neural networks," *Network and Distributed Systems Security Symposium (NDSS)*, 2018, doi: 10.14722/ndss.2018.23198.
- [18] J. Rauber, W. Brendel, and M. Bethge, "Foolbox: a python toolbox to benchmark the robustness of machine learning models," *arXiv preprint arXiv:1707.04131*, 2017.
- [19] J. Kaur, M. A. Khan, M. Iftikhar, M. Imran, and Q. E. U. Haq, "Machine learning techniques for 5G and beyond," *IEEE Access*, vol. 9, pp. 23472–23488, 2021, doi: 10.1109/ACCESS.2021.3051557.
- [20] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: a comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1617–1655, 2016, doi: 10.1109/COMST.2016.2532458.
- [21] V. M. Priya and M. Rajendiran, "A secure framework to improve the channel frequency in mobile ad hoc network," *International Journal of Engineering Research and Technology (IJERT)*, vol. 07, no. 05, May 2018.
- [22] C.-X. Wang, M. Di Renzo, S. Stanczak, S. Wang, and E. G. Larsson, "Artificial intelligence enabled wireless networking for 5G and beyond: recent advances and future challenges," *IEEE Wireless Communications*, vol. 27, no. 1, pp. 16–23, Feb. 2020, doi: 10.1109/MWC.001.1900292.
- [23] B. Ozpoyraz, A. T. Dogukan, Y. Gevez, U. Altun, and E. Basar, "Deep learning-aided 6G wireless networks: a comprehensive survey of revolutionary PHY architectures," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 1749–1809, 2022, doi: 10.1109/OJCOMS.2022.3210648.
- [24] L. Zhang, S. Lambotaran, G. Zheng, G. Liao, B. AsSadhan, and F. Roli, "Attention-based adversarial robust distillation in radio signal classifications for low-power IoT devices," *IEEE Internet of Things Journal*, vol. 10, no. 3, pp. 2646–2657, Feb. 2023, doi: 10.1109/IJOT.2022.3215188.
- [25] O. Ibitoye, R. Abou-Khamis, A. Matrawy, and M. O. Shafiq, "The threat of adversarial attacks on machine learning in network security--a survey," *arXiv preprint arXiv:1911.02621*, 2019.
- [26] X. Yuan, P. He, Q. Zhu, and X. Li, "Adversarial examples: attacks and defenses for deep learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2805–2824, Sep. 2019, doi: 10.1109/TNNLS.2018.2886017.
- [27] J. W. Stokes, D. Wang, M. Marinescu, M. Marino, and B. Bussone, "Attack and defense of dynamic analysis-based, adversarial neural malware detection models," in *MILCOM 2018 - 2018 IEEE Military Communications Conference (MILCOM)*, Oct. 2018, pp. 1–8, doi: 10.1109/MILCOM.2018.8599855.
- [28] F. O. Catak, M. Kuzlu, H. Tang, E. Catak, and Y. Zhao, "Security hardening of intelligent reflecting surfaces against adversarial machine learning attacks," *IEEE Access*, vol. 10, pp. 100267–100275, 2022, doi: 10.1109/ACCESS.2022.3206012.
- [29] B. Peng, B. Peng, J. Zhou, J. Xie, and L. Liu, "Scattering model guided adversarial examples for SAR target recognition: attack and defense," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 2022, doi: 10.1109/TGRS.2022.3213305.

BIOGRAPHIES OF AUTHORS






Rajendiran Muthusamy     is a professor in the Department of Computer Science and Engineering, Panimalar Engineering College, Chennai, India. He holds a Ph.D. degree in computer science and engineering with specialization in mobile ad hoc networks. His research areas are mobile networks, machine learning, data analysis and routing protocol. He can be contacted at email: mrajendiran@panimalar.ac.in.






Charulatha Kannan    received the B.E., and M.E., degrees in computer science and engineering from Panimalar Engineering College, Anna University, Tamil Nadu, India. She is currently working as an assistant professor in the Department of Artificial Intelligence and Data Science, Panimalar Engineering College, Chennai, Tamil Nadu. She can be contacted at email: charulathakannan1971@gmail.com.






Jayarathna Mani    received the B.E., in computer science and engineering from St. Peter's University and M.E., degrees in computer science and engineering from Panimalar Engineering College, Anna University, Tamil Nadu, India. She is currently working as an assistant professor in the Department of Computer Science and Engineering, Panimalar Engineering College, Chennai, Tamil Nadu, India. She can be contacted at email: rathnajaya98@gmail.com.



Dr. Rathinasabapathi Govindharajan    is working as an associate professor and has gained academic qualifications and experience working in Panimalar Engineering College (Autonomous). He has extensive research experience in nano composites, characterization techniques in nano composite materials. He is also involved in the development of automatic wet grinder for home involving automation by implying AI techniques. He can be contacted at email: rajansaba@gmail.com.



Karthikeyan Ayyasamy    earned a bachelor's degree in electronics and communication engineering from IFET College of Engineering, Villupuram, which was affiliated to Anna University, Chennai. He received Master's degrees in Information Technology from Anna University, Regional Campus, Coimbatore. He published 16 research articles in machine learning, information retrieval, and data analytics. He currently serves as the assistant professor, in the Department of Computer Science and Engineering, Panimalar Engineering College, Chennai. He has 15 years of experience in teaching. He can be contacted at email: keyanmailme@gmail.com