Revista Innovación y Software
Vol. 4, No. 2, Mes Septiembre-Febrero, 2023
ISSN: 2708-0935
Pág. 52-62
https://revistas.ulasalle.edu.pe/innosoft

Esta obra está bajo una Licencia Creative Commons Atribución 4.0 Internacional.

Tipo de artículo: Artículos originales
Temática: Inteligencia artificial
Recibido: 25/03/2023 | Aceptado: 08/07/2023 | Publicado: 30/09/2023

Identificadores persistentes:
DOI: 10.48168/innosoft.s12.a108
ARK: ark:/42411/s12/a108
PURL: 42411/s12/a108

Análisis de fugas de datos en redes inalámbricas mediante pruebas supervisadas y no supervisadas

Data leakage analysis in wireless networks using Super-vised and Unsupervised Testing

Shahzad Ashraf [1], Zeeshan Aslam [2]

[1] NFC Institute of Engineering and Technology Multan Pakistan. nfc.ie@hotmail.com
[2] Alfanar Global Development Saudi Arabia. Damman, Arabia Saudita. zeeshan.aslam@alfanar.com

[*] Autor para correspondencia: nfc.ie@hotmail.com

Resumen

Debido al creciente número de espectros inalámbricos, las múltiples frecuencias están enredando el proceso de gestión de recursos, lo que dificulta el funcionamiento. Además, los datos anteriores se vuelven vulnerables cuando se reciben informes de enigma de fuga de datos. En esta situación, es indispensable asegurar los datos en el conjunto de datos y detectar la cantidad real de datos durante el mecanismo de transformación de recursos en redes inalámbricas. Se ha desarrollado un sistema para detectar la fuga de datos utilizando técnicas de prueba supervisadas y no supervisadas mediante simulación en Python. Se obtienen los resultados previstos y reales, que se reducen mediante pruebas supervisadas y no supervisadas, el resultado sigue siendo del 96,03% y 94,53% respectivamente.

Palabras clave: Pruebas supervisadas, pruebas no supervisadas, redes neuronales, redes inalámbricas.

Abstract

Due to an increasing number of wireless spectrums, the multiple frequencies are tangling resource management process that results hindrance in operation. In addition, the previous data become vulnerable when reports are received for data leakage enigma. In this situation, it is indispensable to secure the data in the dataset and detect the actual amount of data during resource transformation mechanism in wireless networks. A system as been developed to detect the leaked data using supervised and unsupervised testing technique by conducting simulation in Python. The targeted and actual outcome is obtained which deduced through supervised and undersized testing, the outcome remained 96.03%, and 94.53% respectively.

Revista Innovación y Software
Vol. 4, No. 2, Mes Septiembre-Febrero, 2023
ISSN: 2708-0935
Pág. 52-62
https://revistas.ulasalle.edu.pe/innosoft

Keywords: *Supervised testing, unsupervised testing, neural network, wireless networks.*

Introduction

With the exponential increase in the number of wireless devices in recent years and due to the rapid growth of wireless services, it has become increasingly important to allocate and manage them accordingly. There is, however, generally a challenge to establish suitable strategies due to the fact that the non-convex objective average rate maximization problem is NP-hard. It is no secret that there has been a growing interest in numerical optimization as it relates to wireless resource allocation in recent years. There are still significant challenges in implementing numerical optimization-based algorithms on practical systems, e.g., high computation costs, despite their ability to solve specific resource management problems with tremendous results [1]

As neural networks (NNs) [2], memorize features of example data during training and unintentionally reveal them during prediction, it is a major concern for machine learning applications to prevent models from revealing sensitive input data details. There is, however, no easy way to accomplish this. The literature is lacking studies that address deep learning-based wireless resource allocation systems with privacy protection.

These radio resource management algorithms cannot always provide an adequate degree of performance due to the dynamic nature of wireless networks. An algorithm that learns from interactions with the environment may be able to better handle such dynamics.

Numerous resource management schemes have been proposed in order to address heterogeneously [3], high service demands, fairness, and starvation, as well as transmission errors due to channel congestion. In spite of this, there are many schemes out there that prioritize voice services while allocating the remaining bandwidth to non-real-time applications without any guarantee that the delay will not affect the voice service.

In wireless communication, when multiple resources exchanges between heterogeneous network environment there is a great chance of distribution of data without the notice of the relevant agencies. This lost of data sometimes create a big hassle and situation becomes aggravate [4]. In this situation, It is imperative to develop a system that should detect the leaked data proactively. After detection of the leaked data the prevention measures can be taken for future. After going through different studies, no specific study is found that can detect the wastage of data during resource transformation mechanism in wireless communication therefore, an intelligent data detection mechanism has been

developed using supervised and unsupervised testing mechanism. This technique utilized hidden layers of NN and then generates the outcome which further processed by conducting simulation in Python.

The following are the main contributions of this work.

- Using intelligent supervised and unsupervised testing, the amount of data leakage would be identified.
- The tested data would be simulated in python to identify the amount of accurate detected data and the amount of lost data.
- The targeted and actual outcome is analyzed to deduce the impact of proposed method.

Rest of the manuscript is arranged as: Section 2 presents a comprehensive overview of previous work. Section 3, is enriched with proposed methodology using supervised and unsupervised testing and Section 4 assesses viability of performance of the proposed method. Finally, Section 5 concludes about findings and future research outlook.

Literature review

It is necessary to advance wireless communication technologies both in terms of scale and complexity to support emerging applications. Machine learning and predictive regression methods [5], have been investigated in order to solve wireless resource allocation problems. Supervised learning involves training neural networks to approximate a function or algorithm so that computation time is minimized. Additionally, their use requires strategies that protect privacy while fulfilling the application requirements.

According to Jorge Cort´es [6], privacy preserving data analysis must take into account dynamic data as well as data exchanged across networks, as well as systems and control perspectives. In order to protect their data against adversaries with arbitrary side information, they adopted differential privacy mechanisms that were initially used to analyze large, static datasets. Under differential privacy constraints, they reviewed how multiple agents can perform signal estimation, consensus, and distributed optimization tasks. Several factors were ignored in this study, including how to deal with signals when multiple spectrums exist and how to avoid tangling them. Further, it is crucial to investigate the appropriate scales for privacy parameters based on specific application domains as well.

Li Ming [7], work was to identify the current situation of small- and medium-sized organizations' human resources, using deep learning data. Through the deep learning approach, human resources can be more productive, and business volumes can be reduced, thereby improving human resource efficiency. His proposed model was improved by implementing a deep neural network. In an analysis of experimental data, several types of decent gradient processes were considered as well as a number of neurons in the hidden layer. Because of the low calculation complexity and fast

training speed, the model fails if there are non-linear relationships between variables, which is common for multiple frequencies.

According to Haoran Sun [8], wireless resource management can be improved by utilizing a learning-based approach. An unknown non-linear map is treated as a resource allocation algorithm and approximated using deep neural networks. As long as a DNN of moderate size is capable of learning accurate and effective non-linear mapping, such a DNN can be used to allocate resources in almost real time, since the input to get the output only involves a few simple operations. In order to approximate some of the algorithms of interest in wireless communications, they developed DNNs to approximate a class of 'learnable algorithms'. Using numerical simulations, the authors demonstrated that DNNs are superior to conventional algorithms for estimating two relatively complex algorithms for power allocation in wireless transmissions. This model only represents a very preliminary step towards understanding the capability of DNN and how to deal with challenging problems such as beamforming for IC/IMAC is still unknown. It also could not answer how to further reduce the computational complexity of DNN?

The application of differential privacy is well documented. However, wireless resource allocation schemes with differential privacy based on Deep Neural Networks (DNNs) [9], have never been studied. This study investigates the impact of Differential Privacy DP) [10], on model convergence and network performance using neural network resource allocation schemes. These research findings show differentially private schemes can produce high-performance models, especially when Convolutional Neural Networks (CNNs) [11], are used.

Proposed method using supervised and unsupervised testing

Differential privacy is achieved by adding noise to the data used in a machine learning model in a way that does not significantly affect the accuracy of the model. This noise ensures that individual data points cannot be easily identified in the model's output. In neural networks, DP can be incorporated into the training process by adding noise to the weights or gradients of the model during training.

There are many functionally interconnected neurons in a neural network that follow a topological structure [12]. The hierarchical structure of neuron is used to categorize neural network models into hierarchical and interconnection models. According to a hierarchical model, neurons are classified into different layers in accordance with their functionality and are interconnected every year. An input, middle, and output layer structure is used in a hierarchical model, while any two neurons are linked in an interconnection model. Hierarchical models are widely used because of their easy analysis and good structure.

3.1 Selection Process

Regression problems include prediction of leaked data during resource allocation process. A regression model is a way of predicting an outcome in machine learning [13]. A regression model may be linear, elastic, neural, polynomial, or ridge. There is a wide range of models which use linear variables, among the most common of which is linear regression. The relationship between the dependent variable and the independent variable is necessary in a linear regression model. This makes linear regression applicable only to problems whose solutions are linearly separable. The model estimates linear relationships between variables because it's simple to calculate and fast to train. Linear regression models are sensitive to outliers, which is a disadvantage.

3.2 Identification of hidden layers

There are several mathematical iterations to calculate amount of actual hidden layers [14]. Two of the most commonly implemented iterations are represented by equations (1) and (2).

$$N_h = \sqrt{n_i \times n_o} \tag{1}$$

$$N_h = \sqrt{n_i \times n_o} + K \tag{2}$$

In these equations, the numbers of input and output layer nodes are represented by n_i and n_o as shown in figure 1. During training, a slight expansion in space is observed, i.e., in the range and the deep neural network should train continuously based on equation (2). 16, 17, and 1 neurons in total were assessed in the input, hidden, and output layers, respectively. The neuron activation function represents sigmoid functions, and the error function is seen to be the quadratic mean square value.

Revista Innovación y Software
Vol. 4, No. 2, Mes Septiembre-Febrero, 2023
ISSN: 2708-0935
Pág. 52-62
https://revistas.ulasalle.edu.pe/innosoft

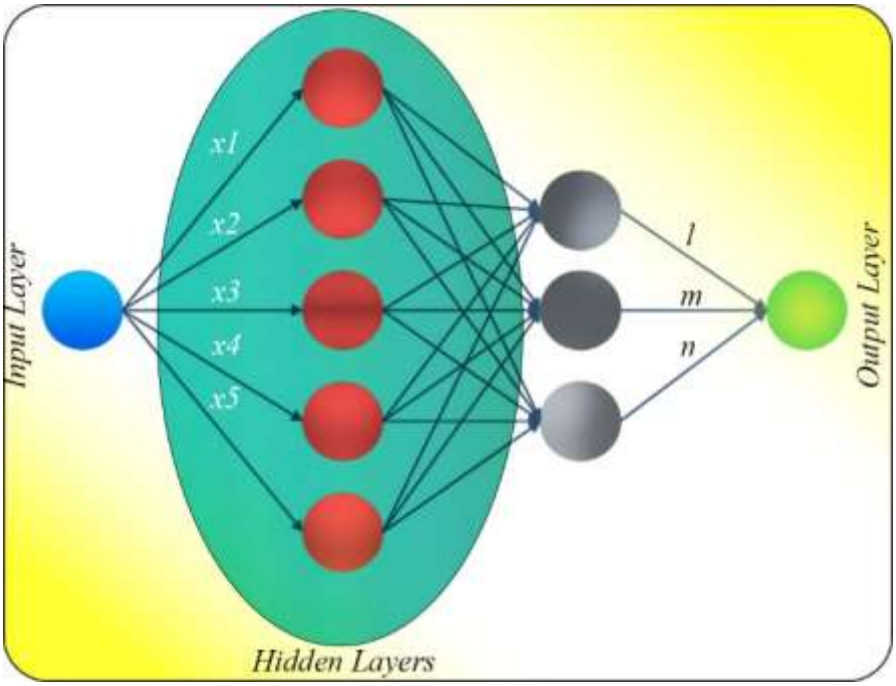

Figure 1. Data leakage detection through Supervised testing mechanism

In supervised testing, n_1 as input layer is applied it generates the sublayers as *x1, x2.....x5*. These layers further undergo for analyze the actual amount of the hidden layers. Iteration by iteration, the hidden layers are identified. These hidden layers actually carries the data that has been separated during resource sharing process. These hidden layers further match the carrying data with sample data and if it matches, the layers change to *l, m,* and *n* in this case. The *l, m*, and *n* layers check the authenticity of the data and save off the ambient data in the form of the noises. Consequently, final data is transferred to the output layer n_o.

In next stage the unsupervised testing is carried out. In unsupervised testing, the data is not labeled and it appears as a mixture of heterogeneous raw table. Figure 2, illustrate the case where at stage one different data has been mixed and ready to transfer to another network. In stage 2, from the mixture of data, some data is labeled as *a1, a2, a3* and *a4* while other data is transferred. The labeled data is screen out from the hidden layers and the statistics of accuracy and the lost, from both supervised and unsupervised technique is placed in table 1 and table 2 respectively. The experiment is performed repeatedly to get all detected data. Based on 750 supervised samples and 250 unsupervised samples, Table 1 illustrates the accuracy values. Observe the final average value of accuracy outcomes after training each hidden layer neuron 25 times consecutively. The execution is terminated once it has run for 500 iterations.

Revista Innovación y Software
Vol. 4, No. 2, Mes Septiembre-Febrero, 2023
ISSN: 2708-0935
Pág. 52-62
https://revistas.ulasalle.edu.pe/innosoft

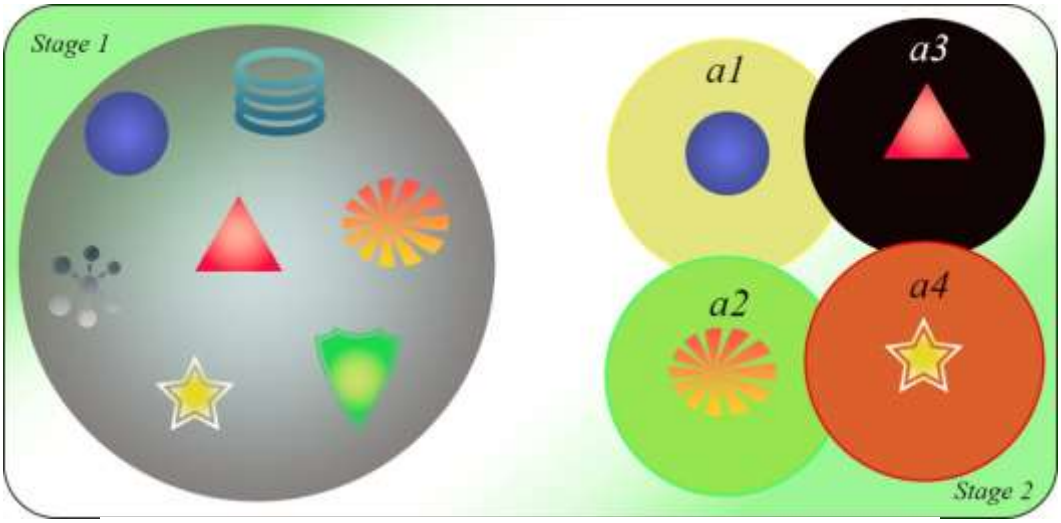

Figure 2. Data leakage detection through Unsupervised testing mechanism

As the number of neurons in the hidden layer increases, both supervised and unsupervised accuracy increase. As shown in table 2, both supervised and unsupervised neural networks are capable of predicting lost data. There is a reduction in loss values with an increase in the number of hidden neurons.

Table 1. Accuracy of leaked data for supervised and unsupervised testing

Hidden layers	Number of Iterations	Accuracy for supervised testing	Accuracy for unsupervised testing
1	30	86.76%	80.67%
2	60	87.62%	80.14%
3	90	88.54%	81.45%
4	120	89.43%	82.65%
5	150	90.67%	83.67%
6	180	90.87%	84.87%
7	210	91.69%	85.95%
8	240	92.78%	86.74%
9	270	92.54%	87.65%
10	300	93.65%	88.62%
11	330	94.75%	89.84%
12	360	94.86%	90.78%
13	390	94.97%	91.62%
14	420	95.46%	92.78%
15	450	95.54%	93.67%

Table 2. Loss of leaked data for supervised and unsupervised testing

Hidden layers	Number of Iterations	Loss supervised testing	Loss for unsupervised testing
1	30	0.0099	0.01017
2	60	0.0098	0.01016
3	90	0.0097	0.01015
4	120	0.0096	0.01014
5	150	0.0095	0.01013
6	180	0.0094	0.01012
7	210	0.0093	0.01011
8	240	0.0092	0.01010
9	270	0.0091	0.01009
10	300	0.0090	0.01008
11	330	0.0089	0.01007
12	360	0.0088	0.01006
13	390	0.0087	0.01005
14	420	0.0086	0.01004
15	450	0.0085	0.01003
16	500	0.0084	0.01002

Performance discourse

After conducting experiments in Python [15], the targeted and the actual generated outcomes are analyzed. The amount of privacy leakage is directly proportional to the batch size and number of hidden layers. To improve the rate, one must raise the batch size and number of hidden layers.

Revista Innovación y Software
Vol. 4, No. 2, Mes Septiembre-Febrero, 2023
ISSN: 2708-0935
Pág. 52-62
https://revistas.ulasalle.edu.pe/innosoft

Figure 3, depicts the curve for targeted and actual outcomes . starting from 20 samples, the target and actual generated outcome is the same, however, both outcome fluctuate ups and down. At sample 70, the targeted detection was aimed at 38, and the same was obtained whereas at samples were reached 120, there was a difference between targeted and actual output. Here the target was 37 but received nearly 41%. Similarly, in sample 220, the target was 70% but the actual outcome was reached to 72%. In the end, the actual outcome left behind the targeted values which was 79% but achieved 92%.

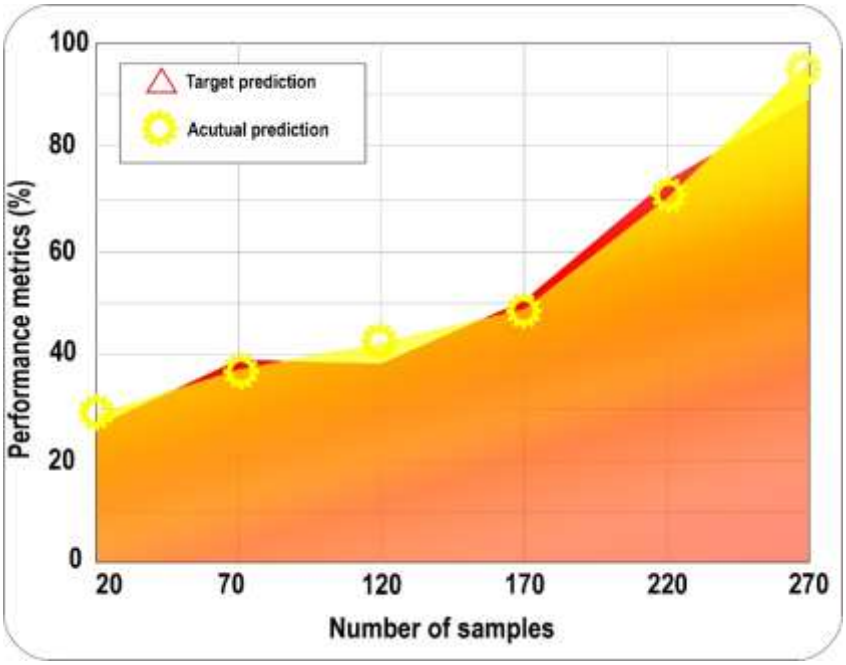

Figure 3. Target versus Actual leaked data analysis

Although the non-normalized dataset [16], may introduce some imperfections, these minute errors are minimized and do not significantly affect the effectiveness of the proposed prediction model. In addition, the proposed method achieved higher accuracy and prediction efficiency as well as a faster convergence rate. Consequently, the prediction model is very accurate. A supervised accuracy of 96.03% was achieved using the proposed method, while an unsupervised accuracy of 94.53% was achieved.

Conclusion

The outcome of this study after analyzing the targeted and actual obtained data showed that using supervised and unsupervised technique to detected the amount of leaked data during data transfer in wireless network is a hallmark of

shrewdness. Models proposed in this study have been shown to perform better in experiments. The proposed method achieved a supervised accuracy of 96.03%, while unsupervised accuracy is remained at 94.53%. Data leakage can be identified using this method with considerable ease.

Contributor Roles

Shahzad Ashraf: Conceptualización, Investigación, Visualización, Metodología, Software, Validación, Redacción - borrador original, Escritura, revisión y edición. **Zeeshan Aslam:** Análisis formal.

References

[1] R. Zeeshan, and A. Muhammad, "Adopting proactive results by developing the Shrewd model of pandemic COVID-19," *Arch. Community Med. Public Health*, vol. 8, no. 2, pp. 062–067, Apr. 2022, doi: 10.17352/2455-5479.000175.

[2] Z. Rasheed, S. Ashraf, N. A. Ibupoto, P. K. Butt, and E. H. Sadiq, "SDS: Scrumptious Dataflow Strategy for IoT Devices in Heterogeneous Network Environment," *Smart Cities*, vol. 5, no. 3, pp. 1115–1128, Sep. 2022, doi: 10.3390/smartcities5030056.

[3] S. Ashraf, "Avoiding Vulnerabilities and Attacks with a Proactive Strategy for Web Applications," vol. 3, no. 2, p. 9.

[4] Z. A. Arfeen, T. Ahmed, S. Ashraf, and S. Saleem, "Succulent link selection strategy for underwater sensor network," *Int. J. Comput. Sci. Math.*, vol. 15, no. 3, p. 224, 2022, doi: 10.1504/IJCSM.2022.10049407.

[5] M. Gao, Z. Chen, H. Naeem, and T. Ahmed, "CED-OR Based Opportunistic Routing Mechanism for Underwater Wireless Sensor Networks," *Wirel. Pers. Commun.*, vol. 125, no. 1, pp. 487–511, Jul. 2022, doi: 10.1007/s11277-022-09561-w.

[6] J. Cortés, G. E. Dullerud, S. Han, J. Le Ny, S. Mitra, and G. J. Pappas, "Differential privacy in control and network systems," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, Dec. 2016, pp. 4252–4272. doi: 10.1109/CDC.2016.7798915.

[7] L. Ming, "A Deep Learning-Based Framework for Human Resource Recommendation," *Wirel. Commun. Mob. Comput.*, vol. 2022, pp. 1–12, Jul. 2022, doi: 10.1155/2022/2377143.

[8] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for wireless resource management," in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Sapporo: IEEE, Jul. 2017, pp. 1–6. doi: 10.1109/SPAWC.2017.8227766.

[9] A. Shahzad, "Towards Shrewd Object Visualization Mechanism," *Trends Comput. Sci. Inf. Technol.*, pp. 097–102, Nov. 2020, doi: 10.17352/tcsit.000030.

[10] D. Muhammad, M. A. Khan, and T. Ahmed, "Fuzzy based efficient Cosmetology Paradigm," vol. 8, pp. 513–520, doi: 10.14741/ijmcr/v.8.4.3.

[11] S. Saleem, and S. Afnan, "FTMCP: Fuzzy based Test Metrics for Cosmetology Paradigm," *Adv. Comput. Intell. Int. J. ACII*, vol. 4, no. 7, pp. 1–13, 2020, doi: 10.5121/acii.2020.7401.

[12] S. Saleem, S. Ashraf, and M. K. Basit, "CMBA - A Candid Multi-Purpose Biometric Approach," *ICTACT J. Image Video Process.*, vol. 11, no. 01, p. 6, 2020, doi: 10.21917/ijivp.2020.0317.

[13] T. Ahmed, "Sagacious Intrusion Detection Strategy in Sensor Network," in *2020 International Conference on UK-China Emerging Technologies (UCET)*, Glasgow, United Kingdom: IEEE, Aug. 2020, pp. 1–4. doi: 10.1109/UCET51115.2020.9205412.

[14] S. Saleem, T. Ahmed, Z. Aslam, and M. Shuaeeb, "Iris and Foot based Sustainable Biometric Identification Approach," in *2020 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, Split, Hvar, Croatia: IEEE, Sep. 2020, pp. 1–6. doi: 10.23919/SoftCOM50211.2020.9238333.

[15] T. Ahmed, Z. Aslam, D. Muhammad, A. Yahya, and M. Shuaeeb, "Depuration based Efficient Coverage Mechanism for Wireless Sensor Network," *J. Electr. Comput. Eng. Innov. JECEI*, vol. 8, no. 2, pp. 145–160, 2020, doi: 10.22061/jecei.2020.6874.344.

[16] A. Yahya *et al.*, "Underwater routing protocols: Analysis of link selection challenges," *AIMS Electron. Electr. Eng.*, vol. 4, no. 3, pp. 234–248, 2020, doi: 10.3934/ElectrEng.2020.3.234.