iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VII **Month of publication:** July 2022

DOI: https://doi.org/10.22214/ijraset.2022.45261

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

A New Technique for Wireless Sensor Network Intrusion Prevention Using Dense Artificial Neural Networks

Saurabh Shrivastava[1], Mr. Gaurav Dubey[2]
[1]*M.Tech Student,* [2]*Assistant professor, Department of Computer Science and Engineering, ITM universe Sitholi Gwalior MP, India,*

Abstract: ***A wide range of industries, including security, healthcare, and industrial settings, currently employ wireless sensor networks (WSNs). With a limited power supply, bandwidth, and energy consumption, WSNs are unique. While traditional networks can be protected in many ways, WSNs cannot be protected in the same way. In order to enhance the service safety of wireless sensor networks, new concepts as well as methodologies were needed. In WSNs, intrusion prevention is the most important issue. For WSN Intrusion Detection (IDS), this research used Dense Artificial Neural Networks (DANN) to develop a DL technique (DeepANN). Compared to the previous models, the proposed Analytical model achieved a 95 percent accuracy rate. The ANN model outperforms the other ML models, with F1 scores of 99, 98, and 96.***
Keywords: Deep Learning, Wireless Sensor Networks, Intrusion Detection Systems, Artificial Neural Network.

I. INTRODUCTION

Computer systems are monitored by a malware detection (IDS), which is a piece of hardware or software that identifies infiltration trends, including such unauthorised access or authorised access that goes beyond the scope of authority. Multiple analyses of the same data are possible when using an IDS because it examines a wide range of CS data sources[1]. The first method involves comparing the data to numerous databases that contain known attacks; each indicator should point to an attempt to circumvent the safeguards that are in place. The second approach is to investigate whether there are any issues with a consumer who is attempting to exceed their restrictions[2]. An intrusion detection system is a device or programme which has the same abilities as an intrusion detection system (IDS), but also has the ability to prevent possible attacks. This device or programme is sometimes referred to as an application[3]. In the modern world, combating cybercrime requires employing methods that are both efficient at identification and effective at preventing it. Examples of common security measures include firewalls, viruses, and a variety of other preventative safeguards and protections. The four primary pillars that support computer security are authentication, confidentiality, accessibility, and authenticity of all data that has been specified[4].

An unusual behaviour, also known as an anomaly, is a pattern of action that occurs within a system that is genderqueer, unexpected, or at random. Because anomalies can have a significant effect on the dependability of a network's operations, ignoring them can have severe repercussions if the network is not properly monitored. When anomalies in a system go unnoticed, they have the potential to expose confidential information to outside world, cause economic stress and/or losses, as well as, even worse, lead to decisions that could have fatal consequences. Although anomalous behaviours are typically associated with security flaws, they can just as easily be brought on by everyday problems with the system's hardware or software. In any case, anomalies in a computer network need to be identified, located, and reduced as quickly as is humanly possible[5].

The detection of anomalies has traditionally been carried out using a statistical foundation. The development of machine learning (ML) has resulted in the availability of new options for locating outliers as a direct consequence of the availability of enormous quantities of data to train complex learning models. This is an intriguing idea, particularly in fields such as the Iot devices, where the emergence of new data patterns makes the application of static models challenging[6].

II. RELATED WORK

Samir Ifzarne et al. present a method of intelligent ID based on incremental machine learning (ML). The model uses a cluster WSN network topology to detect incursions and classify them in real time. To detect intrusions in a timely and efficient manner, the ID-GOPA model is proposed. As a feature selection, it made use of the gain ratio to minimise both the characteristics and the processing load.

Feature selection is a critical part of improving the algorithm's performance when using passive-aggressive methods as an incremental learning machine. This shows that the model is highly accurate, as the simulation results are 96% accurate compared to the offline models. Since it can be used for any purpose, the model is superior to previous systems[7]. In this work, Francesco Cauteruccio how short and long term algorithms can reliably identify anomalies in a WSN are examined. Locally identified 785 potential irregularities using the short-term approach and highlighted time periods of potential importance that were then transferred to a cloud storage service for longer-term analysis were successfully identified. Long-term techniques are useful for identifying anomalous periods in time. False positives, excessive processing time, and other drawbacks can be reduced while advantages like speed and accuracy are improved by combining short-term and long-term strategies[8]. Xianhao Shen et alresearch .'s in WSN, the challenge of detecting anomalies in data Marked mode as well as neural network based structure are used to develop the CNN model for anomalous data. In the studies, they used DA, TPR, and PRE to evaluate the performance of three new network models as well as compared them to the a previous cart model. Three models are described in this paper, with the M2 model performing the best in experiments[9]. Anton Kanev proposes an ANN-based[10] Network anomaly detection method for "smart home" systems. "Smart home" and building automation systems have never before been tested with a hybrid ANN technique for anomaly detection. This is a scientific first." The results of the experiment show that the ANN outperforms traditional ML methods, with an AUC of 0.9689[10].

III. METHODOLOGY

The proposed model used the CIC KDD NSl[12] dataset from (https://www.unb.ca/cic/datasets/nsl.html), which contains two files for training and testing machine learning algorithms: KDDTrain.txt and KDDTest.txt. Finally, we turned the.txt data into a pandas data frame, which contains 42 characteristics containing numerical and textual data. After that, we used the describe() method to compute and display summary statistics for a Python data frame in terms of count, mean, standard deviation, minima, and maxima, which has been giving statistical information. It also works with Pandas series objects and data frame columns. Count the number of different attacks in the data frame. For visualization, data were analysed using univariate, bivariate, and multivariate features for analytical attacks (normal, dos, probe, r2l, and u2r) on various networking payload protocols such as TCP, UDP, and ICMP. For density, we used a bar graph analysis of the label, protocol, flag, and duration graphs.

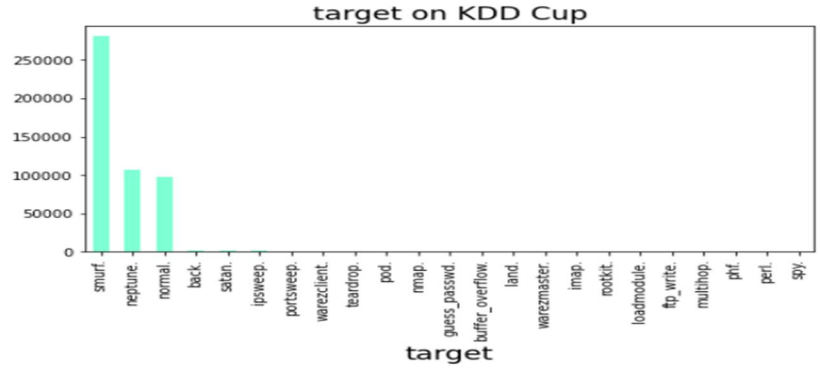


Figure 1: Bar Plot of Sub Attacks.

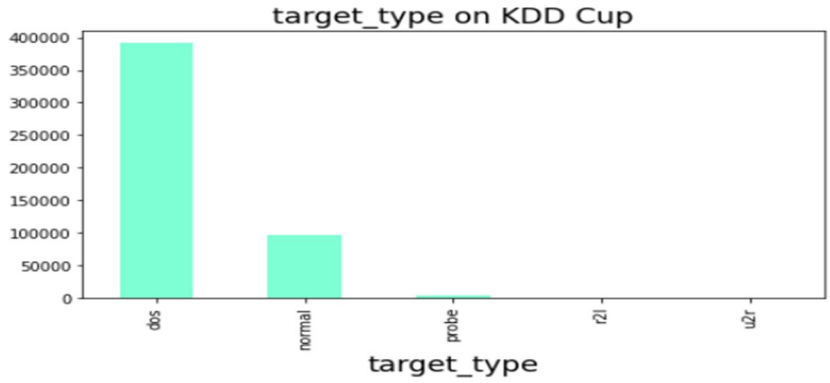


Figure 2 Bar Plot of 5 Main Attacks.

Pictured here is a multivariate graph evaluation of the label. There are labels for each type of attack, such as "normal," "nNeptune devil," "ipsweep port sweep," and so on. For each label in figure 2, a bar graph analysis demonstrates the amount of count data for the various attack kinds: regular; Dos; Probe; R2L; and U2R.

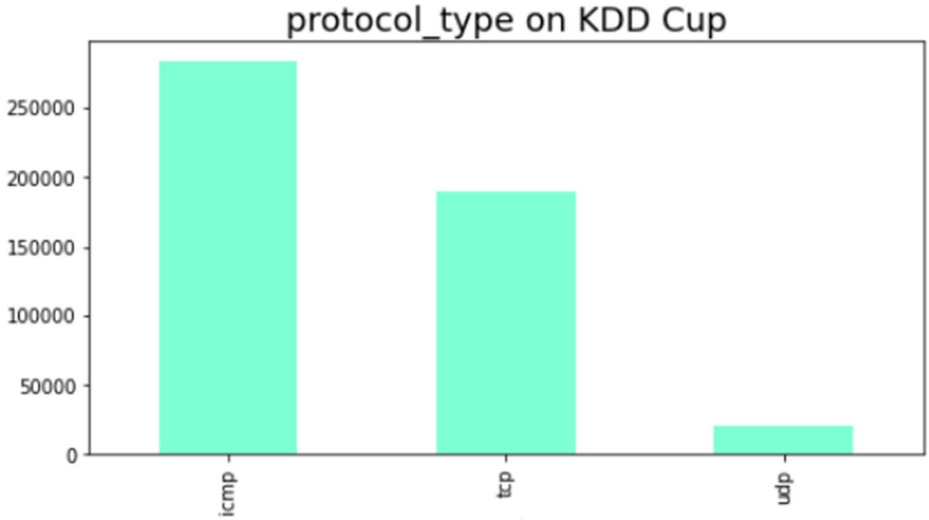


Figure 3: Plot of Protocols Used.

An analysis from several numbers for different network payload procedures such as TCP, UDP and ICMP is shown in Figure 3. Following the binning of data, the bandwidth duration histogram is shown in Figure 4. The image displayed a bar graph showing the total number of flags in the area. SF, S0, REJ, RSTR, RSTO, and other flag counts were shown. It was decided to use bivariate analysis to determine the types of attacks and how they were carried out on the datasets. A dataset's size, as well as the type of attack and its associated services. The 1323 matrix yielded this result.

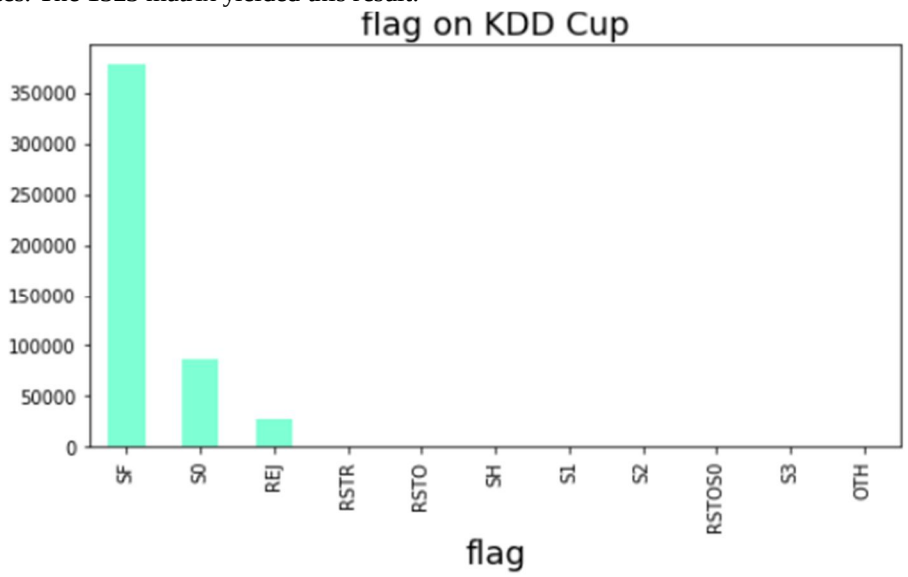


Figure 4: Flag use .

We prepared data for training as well as testing for consists of multi classification using Label and One-Hot encoding[11] Sk-learn is an effective method for converting the numerical value of a feature's category level into a numerical value. n classes-1 is the maximum number of values that Label-Encoder can encode for each of the labels in a file. Categorical variables can be encoded as binary vectors, for example. Begin by converting the categories into integers. Once this is done, a binary vector with all zeros except for the integer's index is generated and labelled with a 1.

Feature selection[12], Afterwards, we decided to use the "intrusion" function. Selecting the features which have the greatest impact on the output variables as well as output that you care about can be done automatically or manually. If your data contains irrelevant attributes, your model's accuracy could be lowered and it could be forced to train on unlabeled data, which would lead to a failure. As it turned out, the X and Y shapes were 125973, 42 and (125973, 42). (125973, 1). To measure the significance of a character, a correlation is performed[13]. Samples that used TCP and resulted in an intrusion, as well as samples that used TCP but didn't lead to an intrusion, are shown in the dataset[14] type and led to the normal situation.

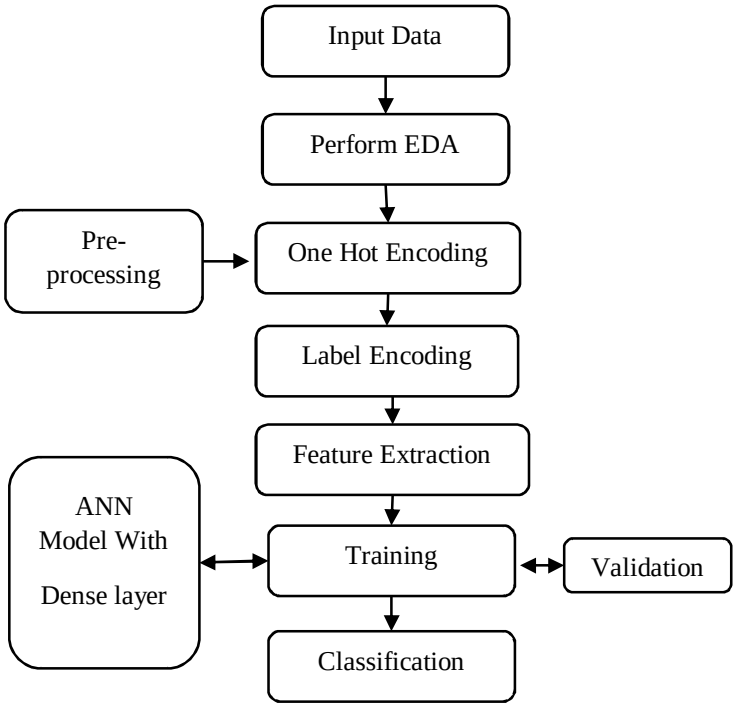


Figure 5 – Proposed Flowchart

To develop a binary classification model, the Deep ANN Classifier is utilised. 29 neurons in the Dense layer are used to train the sequential model, and the Softmax activation function is used to complete the output layer. The ADAM optimizer is used to train the model. Table 1 lists the various parameters that are used to train the model.

Table 1: Parameters used for Training

Model	Sequential
Neural Network	Artificial Dense Network Kears
Training data	80%
Testing data	20%
Training Data	(81581,42)
Test Data	(9072,42)
Validation Data	(9070,42)
Epoch	50
Layers	DENSE with 29 neurons
Output layer	5 for Classes
Important Feature Selection	RandomForest
Output function	SOFTMAX
Loss function	Sparse categorical cross-entropy
Optimizer	ADAM
Learning Rate	0.001
Metrics	ACCURACY

As a result of the model training, the accuracy vs the epoch graph and the loss versus epoch graph plot for the training and testing sets datasets are shown in Figure 6.

Table 3 shows that the proposed ANN model performs better the other ML methods. The ANN model has a precision of 96.38 percent, a recall of 98.94 percent, and an F1 score of 97.64 percent accuracy of 94.45 percent. Other machine learning models, such as ANNs, were compared to our model's results[15], Decision Tree, Support Vector Machine[16], etc.

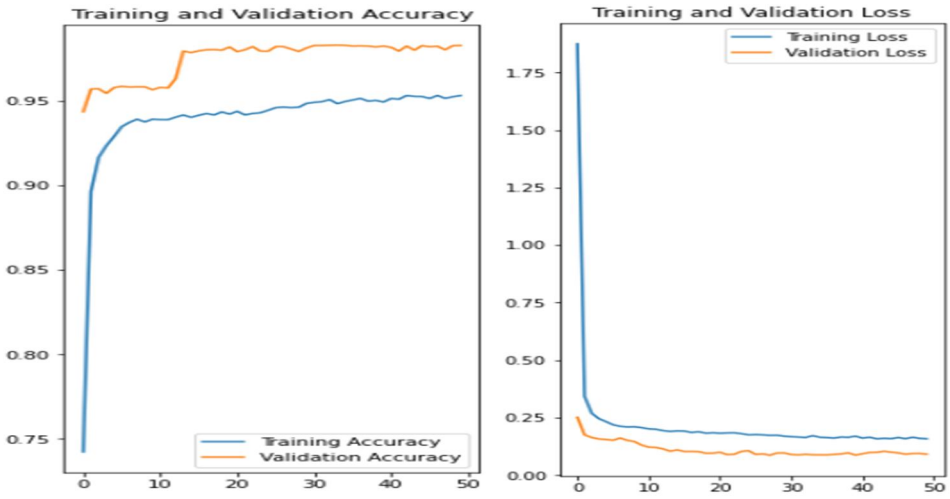


Figure 6: Accuracy and Loss Graph of Train and Test Data

Among the performance metrics used to gauge the final result are accuracy, precision, recall, and the F-measure.

1) *True Negatives (TN):* Negative values (meaning the real class value is 'no' as well as the anticipated class value is 'not') are correct. This is the case, for example, if an actual class implies that this passenger has died and the expected class confirms that this is the case. False positive and false negatives are in opposition to each other when the real class is discovered.
2) *True Positives (TP):* The predicted and actual class values are both correct. Consider the distinction between the "this passenger has survived" actual class value and the "this rider is likely to be same one next time" predictive class value.
3) *False Negatives (FN):* True for the actual class, but false for the predicted one. In other words, the value of each class of passengers may reveal whether or not those passengers have survived.
4) *False Positives (FP):* For example, if the class value is "No," but the class value is "Yes," then this person has died.
5) *Accuracy:* As the number of measurements increases, so does the basic accuracy metric. The statistical precision of symmetrical datasets is improved because the number of false negatives and false positives is nearly equal. Accuracy = TP+TN/TP+FP+FN+TN
6) *Precision:* It's the percentage of predicted positive events that were actually observed. Precision = TP/TP+FP
7) *Recall (Sensitivity):* Based on everything that was actually observed in class, it's a ratio of positive comments that's been precisely anticipated. Recall = TP/TP+FN
8) *F1 Score:* Weighted base percentage of precision and recall is used to calculate this result. False negatives and positives are accounted for in this score.

Table 2: Model Matrices and Performance Evaluation.

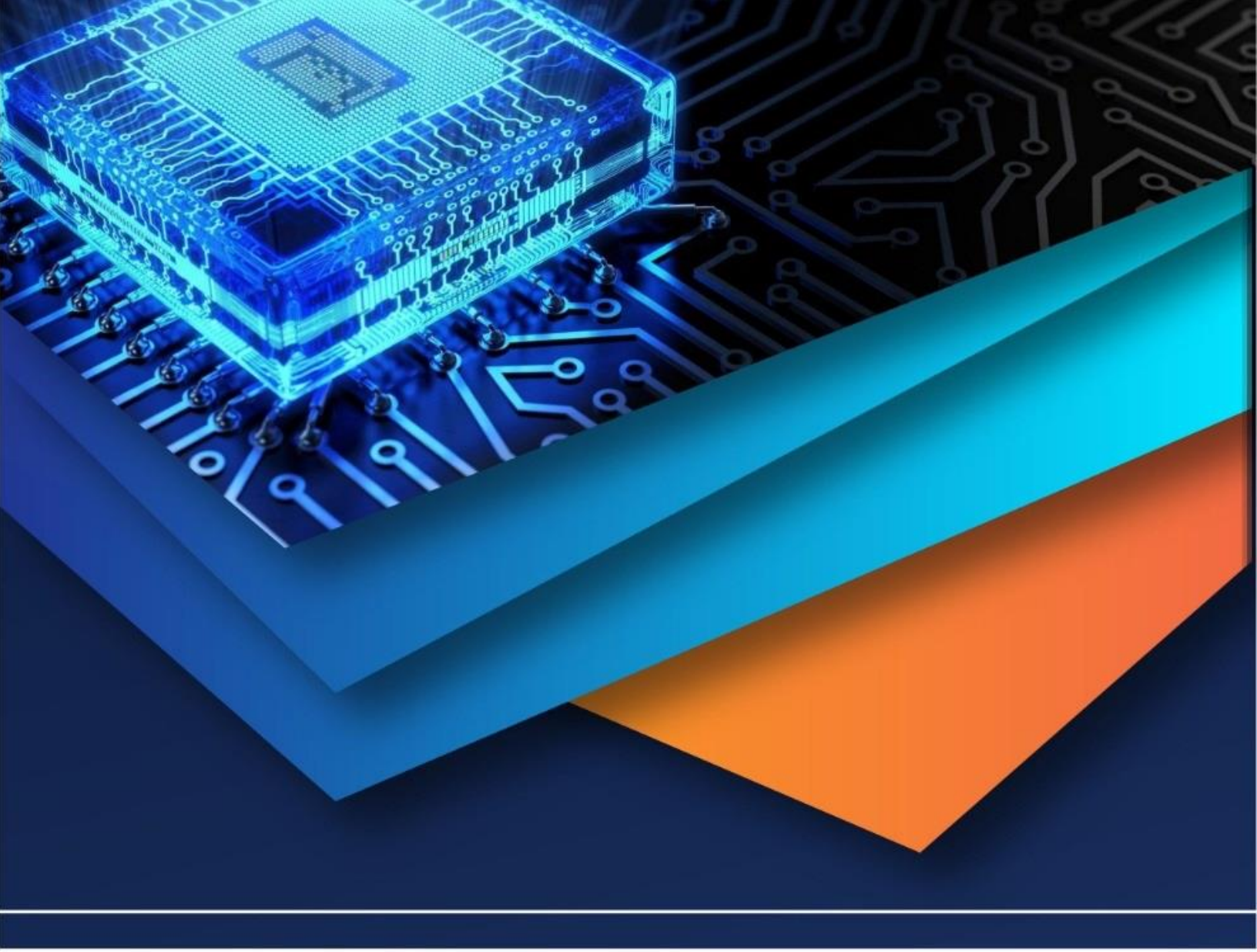
Model	Accuracy	Precision	Recall	F1-score
Propose Dense-ANN	95	99	98	96
DT	88	85	94	89
SVM	85	85	87	86
RF	74	81	74	70
XGBoost	77	81	77	73
LSTM	78	78	78	75

IV. CONCLUSION

WSN's practical applications face a significant identification challenge. As the service area grows and the volume of data increases, the threat as well as consequences of network attacks in WSN cannot be ignored.. If an ID system can only handle certain types of attacks, it's useless in the case of unforeseen threats. In the WSN, relying on an IDS that can clearly detect attacks is difficult. Machine learning is used to develop a smart ID strategy in this study. Attacks are detected using a grouped WSN network topology, which is used to classify them.

REFERENCES

[1] G. Divyashree, A. Durgabhavani, M. Kavya, A. Gudoor, and M. B. Shetty, "Intrusion detection system in wireless sensor network," Int. J. Recent Technol. Eng., vol. 8, no. 1, pp. 2047–2051, 2019.
[2] P. R. Chandre, P. N. Mahalle, and G. R. Shinde, "Machine Learning Based Novel Approach for Intrusion Detection and Prevention System: A Tool Based Verification," Proc. - 2018 IEEE Glob. Conf. Wirel. Comput. Networking, GCWCN 2018, pp. 135–140, 2019, doi: 10.1109/GCWCN.2018.8668618.
[3] L. Sheeba and V. . Meenakshi, "A Brief survey on Intrusion Detection System for WSN," Int. J. Comput. Trends Technol., vol. 40, no. 3, pp. 109–113, 2016, doi: 10.14445/22312803/ijctt-v40p121.
[4] A. H. Farooqi and F. A. Khan, "A survey of intrusion detection systems for wireless sensor networks," Int. J. Ad Hoc Ubiquitous Comput., vol. 9, no. 2, pp. 69–83, 2012, doi: 10.1504/IJAHUC.2012.045549.
[5] S. H. A. H. Baddar, A. Merlo, and M. Migliardi, "Anomaly detection in computer networks: A state-of-the-art review," J. Wirel. Mob. Networks, Ubiquitous Comput. Dependable Appl., vol. 5, no. 4, pp. 29–64, 2014, doi: 10.22667/JOWUA.2014.12.31.029.
[6] I. Butun, S. D. Morgera, and R. Sankar, "A survey of intrusion detection systems in wireless sensor networks," IEEE Commun. Surv. Tutorials, vol. 16, no. 1, pp. 266–282, 2014, doi: 10.1109/SURV.2013.050113.00191.
[7] S. Ifzarne, H. Tabbaa, I. Hafidi, and N. Lamghari, "Anomaly Detection using Machine Learning Techniques in Wireless Sensor Networks," J. Phys. Conf. Ser., vol. 1743, no. 1, 2021, doi: 10.1088/1742-6596/1743/1/012021.
[8] F. Cauteruccio et al., "Short-long term anomaly detection in wireless sensor networks based on machine learning and multi-parameterized edit distance," Inf. Fusion, vol. 52, no. June 2018, pp. 13–30, 2019, doi: 10.1016/j.inffus.2018.11.010.
[9] "A Method for Detecting Abnormal Data of Network Nodes Based on Convolutional Neural Network," pp. 1–12.
[10] A. Kanev et al., "Anomaly detection in wireless sensor network of the 'smart home' system," Conf. Open Innov. Assoc. Fruct, vol. 2017-April, pp. 118–124, 2017, doi: 10.23919/FRUCT.2017.8071301.
[11] P. Vadapalli, "Label Encoder vs One Hot Encoder in Machine Learning," upGrad, 2021. .
[12] Rahul Bajaj, "Feature Selection Techniques in Machine Learning," Geeks for Geeks, 2021. .
[13] A. Upadhyay, "What Is Correlation in Machine Learning?," Analytics Vidhya, 2020. .
[14] JayGala, "What is Transmission Control Protocol (TCP)?," Geeks for Geeks, 2021.
[15] M. Gupta, "ML | Linear Regression," Geeks for Geeks, 2018. .
[16] S. Morris, "Image classification using SVM," Rpubs.Com, 2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY