Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 13 Issue 1, January - 2026

Evaluation of Decision Tree Binary classifier for IIoT networks intrusion detection application

Michael Oghale Ighofiomoni
Department of Computer Engineering
Southern Delta University, Ozoro, Delta State, Nigeria
Ighofiomonimo@dsust.edu.ng

Abstract

Decision Tree Binary classifier (DTBC) for IIoT networks intrusion detection application is presented. The study focused on examining the impact of data balancing on the performance of the classifier model by applying the DTBC in the binary classification of packets in the case study Industrial Internet of Things Dataset (IIoTID) into Normal Packet (NP) and Attack Packet (AP) using the balanced dataset and also using the imbalanced dataset. The data balancing was done using Synthetic Minority Over-sampling Technique (SMOTE) which increased the overall data samples in the balanced dataset. The results showed that without data balancing, the attack class has precision of 91.55% whereas the normal class has precision of 99.31 % which is 7.78 % ahead of the attack class. Notably, the attack class is 44% while the normal class is 56% in the dataset. As such, the model performance favors the majority class which is the normal class. After data balancing, the attack class has precision of 99.6% which is about 8.07% improvement. Again, the balanced dataset case has about 8.16% improvement in the overall model precision, showing that data balancing has significant effect on the classifier performance. Also, without data balancing, the Accuracy is 93.65 % whereas the balanced data case had Accuracy of 99.73 % which is 6.08 % improvement. The ROC AUC was 92.15 % whereas the balanced data case had ROC AUC of 99.73 % which is 7.58 % improvement. Furthermore, without data balancing, the training time was 29.92 seconds whereas the balanced data case had training time of 31.61 seconds which is 8.51 % improvement. This is because the balanced dataset model was trained with higher number of data samples than the case of imbalanced dataset. In all, the study showed that data balancing improved the overall performance of the classifier model.

Keyword: Data Balancing, Decision Tree Binary classifier, Industrial Internet of Things Dataset (IIoTID), Synthetic Minority Over-sampling Technique (SMOTE), Networks Intrusion Detection

1. Introduction

In recent years, there has been growing application of Internet of Things (IoT) technologies in various industries [1,2,3]. Such IoT applicable in the industries are known as Industrial Internet of Things (IIoT) [4]. It facilitates more advanced automation of industrial process, and greater utilization of industrial robots which minimizes humane involvement in the industrial process [5,6,7].

In any case, one of the major drawbacks of the IIoT applications is cyber threats which tends to undermine he gains of IIoT [8,9,10]. As such, efforts are constantly being made to mitigate the attacks by implementing IIoT intrusion detection. In this work, a Decision Tree Binary Classifier (DTBC) model is presented for binary classification of IIoT traffic into normal packet and attack packet [11,12]. This is essential in the intrusion detection module to identify attack class and then take necessary measures to deny them access to the network.

One of the issues associated with classification model is imbalanced dataset; in most cases the attack class is in the minority while the normal packets form the majority class [13,14]. Such imbalanced distribution of data samples affect the performance of the classifier models in identifying the attack packets [15,16]. Accordingly, in this work, the DCBC is evaluated using the balanced and the imbalanced datasets and the performance of the classifier in the two cases are compared to ascertain the effect of dataset balancing on the performance of the model.

2. Methodology

The work aims at using Binary Tree Classifier (DTBC) model to categorize the traffic in Industrial Internet of Things (IIoT) into Normal Packet (NP) and Attack Packet (AP). Specifically, the study examined the classification using the imbalanced dataset and the balanced dataset of the IIoT network. The essence if to identify the impact of the data balancing of the performance of the Decision **Tree** binary classifier model.

2.1 Decision Binary Tree Classifier (DTBC) Model

The Decision **Binary Tree Classifier (DTBC) model** is typically a Decision Tree Model (DTM) that is used to classify data into two classes; in this case it is about classifying the data packets into normal packets and attack packets.

Decision Tree Model (DTM), with the architecture in Figure 1, is non-parametric supervised learning algorithm that performs classification by partitioning the dataset into smaller and smaller subsets with corresponding creation of decision trees that will use the data subset for decision making.

The outcome of the two actions lead to the creation of a tree that has nodes for decision making where each node has one more leaf nodes. It is chosen for its interpretability and speed.

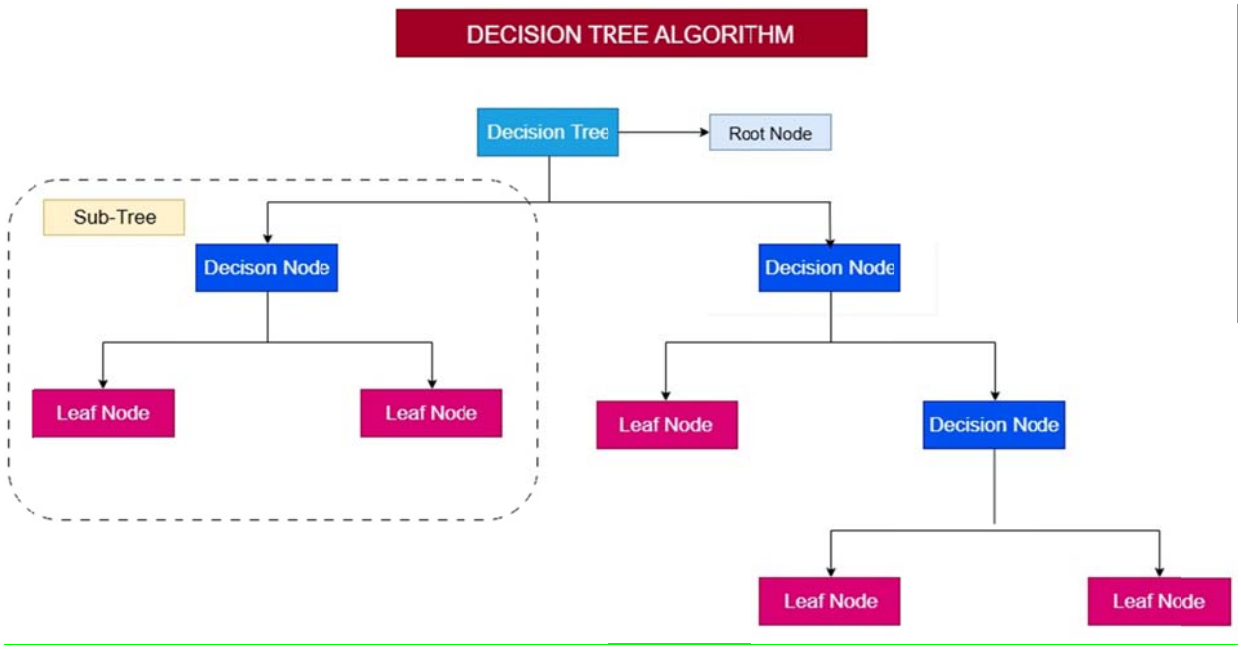

Figure 1 The Architecture of Decision Tree Model

Basically, the DTM works like a flowchart, splitting the data based on values of specific features. For example, in this research, the DTM may ask: "Is Duration > 5?"; based on the answer, the data is split and the process continues until a final decision (Normal or Attack) is made at a leaf node. In the DTM implementation in the research, all the 79 features are available, and the model selects the best ones to split on at each point. Features that better separate the classes are prioritized higher in the tree. For each of the data points, the model follows the decision path down to a leaf node that determines whether it is classified as Normal or Attack. The key parameter settings used for the DTM are as follows;

(i) `random_state=42` → ensures reproducibility

(ii) `criterion='gini'` → evaluates the quality of splits

(iii) `max_depth=None` → tree can grow until pure leaves or minimum samples reached

(iv) `min_samples_split=2`, `min_samples_leaf=1` → default settings for node splits

2.2 The Description of the Study Dataset

The study used the Industrial Internet of Things Dataset (IIoTID) comprising 820,834 rows and 68 columns (or 86 features). The dataset is imbalanced with 234,220 Normal Packet (NP) data samples and 180,337 Attack Packet (AP) data samples (as shown in Table 1 and Figure 2). The Synthetic Minority Over-sampling Technique (SMOTE) data balancing was applied [17] and the AP class up-sampled to equal the NP class data samples, each with 234,220 data samples (as shown in Table 1 and in Figure 3).

Table 1 The distribution of the data samples between the two data classes in the dataset

	Normal Packet (NP)	Attack Packet (AP)	Total
Number of Data Samples Before SMOTE Data Balancing	234,220	180,337	414,557
Number of Data Samples After SMOTE Data Balancing	234,220	234,220	468,440

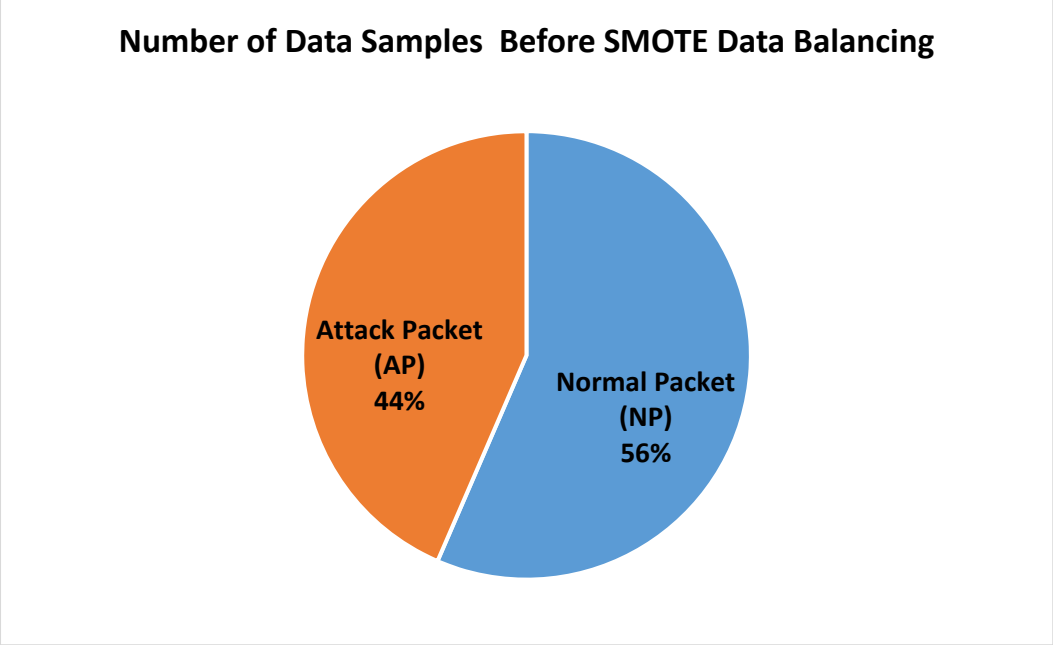

Figure 2 Number of Data Samples before SMOTE Data Balancing

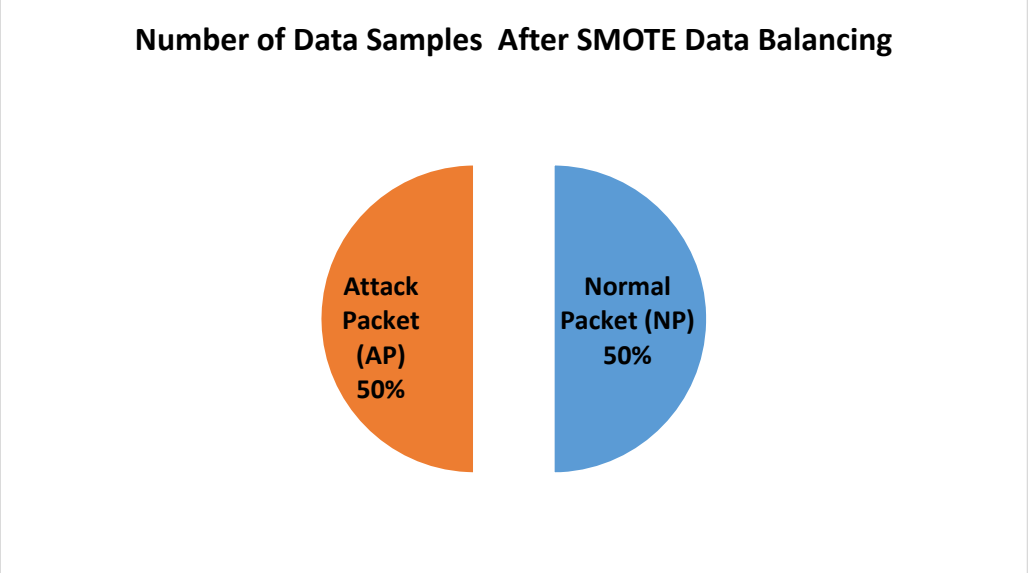

Figure 3 Number of Data Samples Before SMOTE Data Balancing

The distribution of the data samples between the training and validation datasets before and after data balancing is presented Table 2 and Figure 4. Accordingly, data splitting of 75% to 25% for training and validation respectively is used.

Table 2 The distribution of the data samples between the training and validation datasets before and after data balancing

	Training Dataset	Validation Dataset	Total
Number of Data Samples for Training and Validation Set Before Data Balancing	310,917	103,639	414,557
Number of Data Samples for Training and Validation Set After Data Balancing	351,330	117,110	468,440

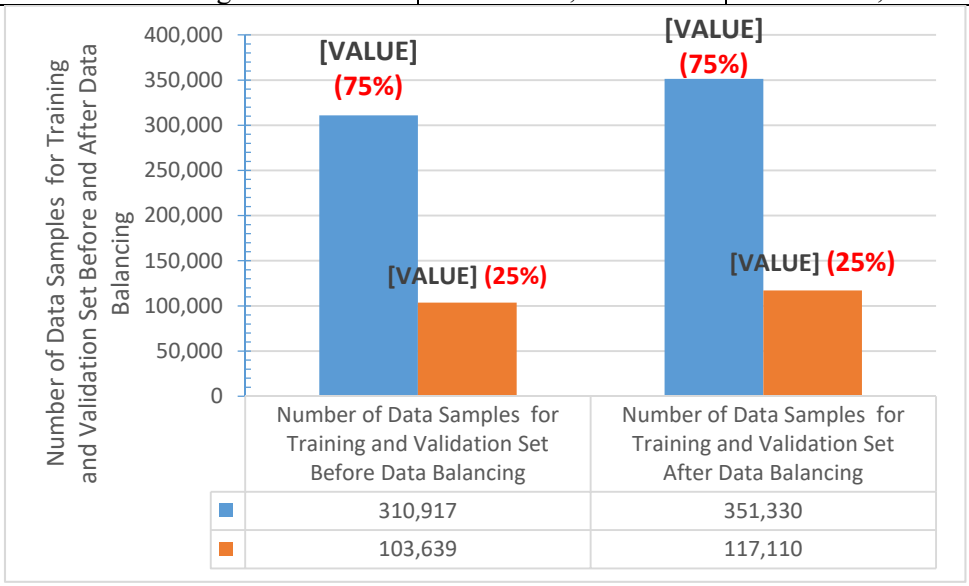

Figure 4 Number of Data Samples for Training and Validation Set Before and After Data Balancing

3. Results and Discussion

The Binary Tree Classifier (DTBC) model was trained and validated using the balanced and unbalanced Industrial Internet of Things Dataset (IIoTID) and the results in terms of the various performance parameters (including precision, Recall, F1-Score, Accuracy, ROC AUC (Receiver Operating Characteristic - Area Under the Curve) and Training Time are presented in Table 3 to Table 5, as well as in Figure 5 to Figure 10.

As shown in Table 3 and Figure 5, without data balancing, the attack class has precision of 91.55% whereas the normal class has precision of 99.31 % which is 7.78 % ahead of the attack class. Notably, the attack class is 44% while the normal class is 56% in the dataset. As such, the model performance favors the majority class which is the normal class. After data balancing, the attack class has precision of 99.6% which is about 8.07% improvement. Again, the balanced dataset case has about 8.16% improvement in the overall model precision, showing that data balancing has significant effect on the classifier performance.

Table 3 The Precision performance of the DTBC model for the balanced and the imbalanced dataset cases

Class	Precision For Balanced Dataset	Precision For Unbalanced Dataset	Difference (%)
Attack	99.62%	91.55%	8.07%
Normal	99.81%	99.31%	0.50%
Macro Avg	99.71%	91.55%	8.16%

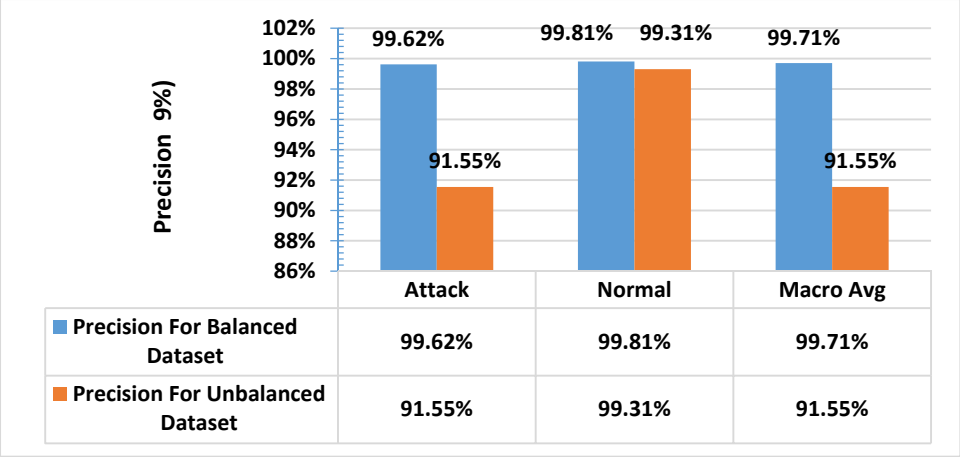

Figure 5 Comparison of the Precision performance of the DTBC model for the balanced and the imbalanced dataset cases

As shown in Table 4 and Figure 6, without data balancing, the attack class has Recall of 98.45% whereas the normal class has Recall of 98.90 % which is 0.45 % ahead of the attack class. After data balancing, the attack class has Recall of 99.75% which is about 1.30% improvement. Again, the balanced dataset case has about 5.39% improvement in the overall model Recall, showing that data balancing has significant effect on the classifier performance.

Table 4 The **Recall performance of the DTBC model for the balanced and the imbalanced dataset cases**

Class	Recall For Balanced Dataset	Recall For Unbalanced Dataset	Difference (%)
Attack	99.75%	98.45%	1.30%
Normal	99.70%	98.90%	0.80%
Macro Avg	99.73%	94.34%	5.39%

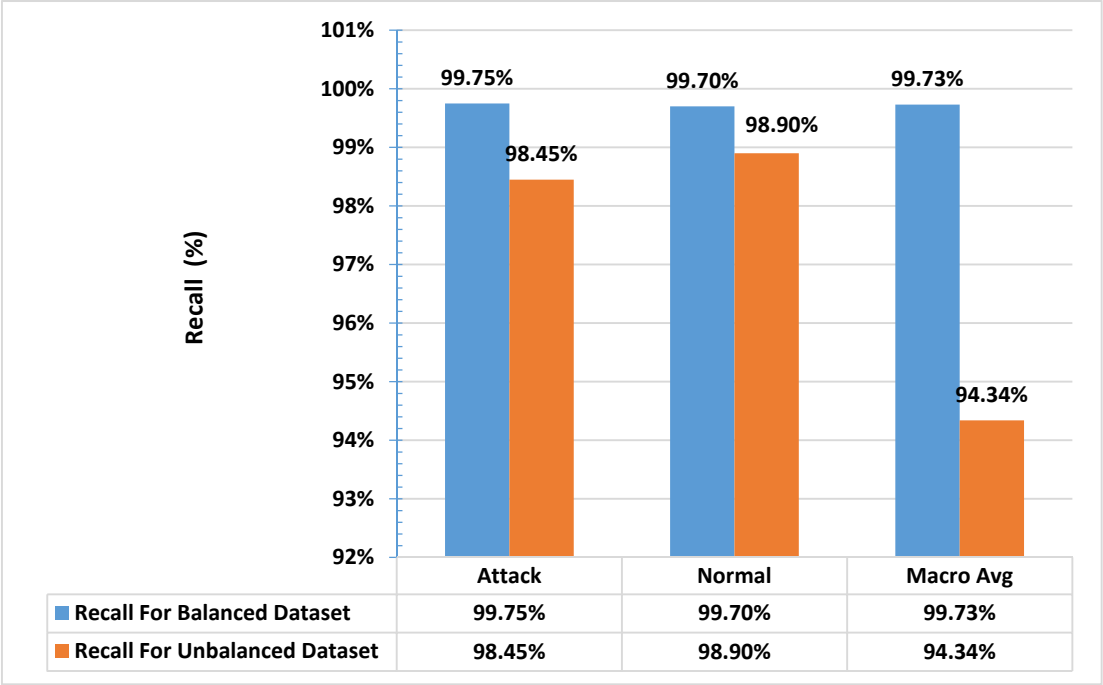

Figure 6 Comparison of the Recall performance of the DTBC model for the balanced and the imbalanced dataset cases

As shown in Table 5 and Figure 7, without data balancing, the attack class has F1 -Score of 99.11% whereas the normal class has F1 -Score of 99.75 % which is 4.23 % ahead of the attack class. After data balancing, the attack class has F1-Score of 99.68% which is about 4.80% improvement. Again,

the balanced dataset case has about 6.79% improvement in the overall model F1 -Score, showing that data balancing has significant effect on the classifier performance.

Table 5 The F1-Score performance of the DTBC model for the balanced and the imbalanced dataset cases

Class	F1-Score For Balanced Dataset	F1-Score For Unbalanced Dataset	Difference (%)
Attack	99.68%	94.88%	4.80%
Normal	99.75%	99.11%	0.64%
Macro Avg	99.72%	92.93%	6.79%

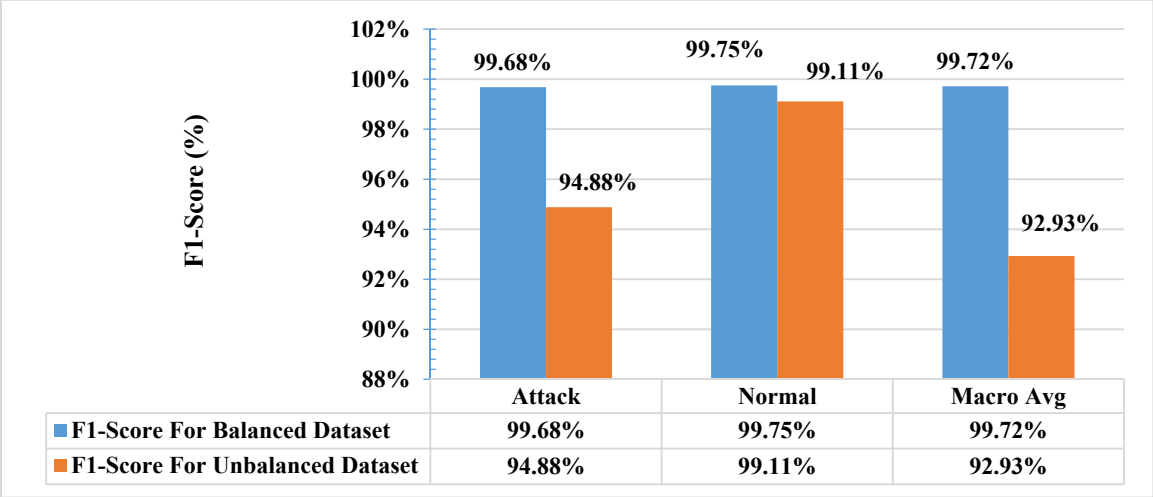

Figure 7 Comparison of the F1-Score performance of the DTBC model for the balanced and the imbalanced dataset cases

As shown in Figure 8, without data balancing, the Accuracy is 93.65 % whereas the balanced data case has Accuracy of 99.73 % which is 6.08 % improvement. This shows that data balancing has significant effect on the classifier performance.

Also, as shown in Figure 9, without data balancing, the ROC AUC is 92.15 % whereas the balanced data case has ROC AUC of 99.73 % which is 7.58 % improvement. This shows that data balancing has significant effect on the classifier performance.

Furthermore, as shown in Figure 10, without data balancing, the training time is 29.92 seconds whereas the balanced data case has training time of 31.61 seconds which is 8.51 % improvement. This is because the balanced dataset model was trained with higher number of data samples than the case of imbalanced dataset.

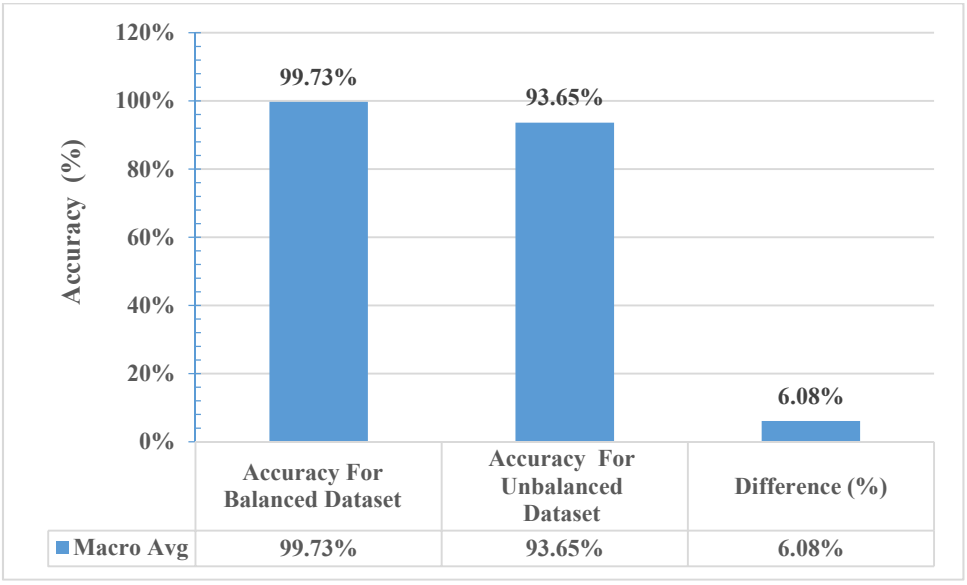

Figure 8 Comparison of the Accuracy performance of the DTBC model for the balanced and the imbalanced dataset cases

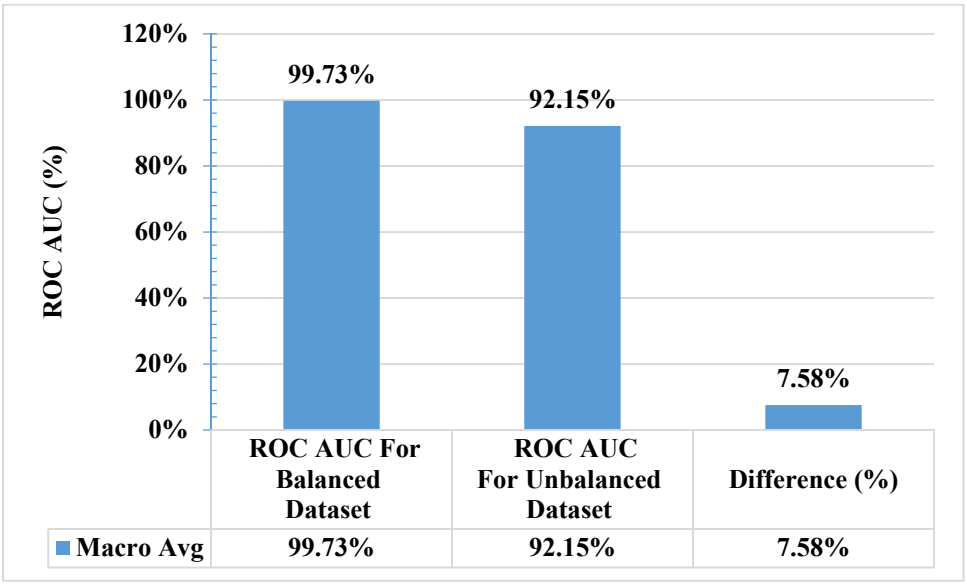

Figure 9 Comparison of the ROC AUC performance of the DTBC model for the balanced and the imbalanced dataset cases

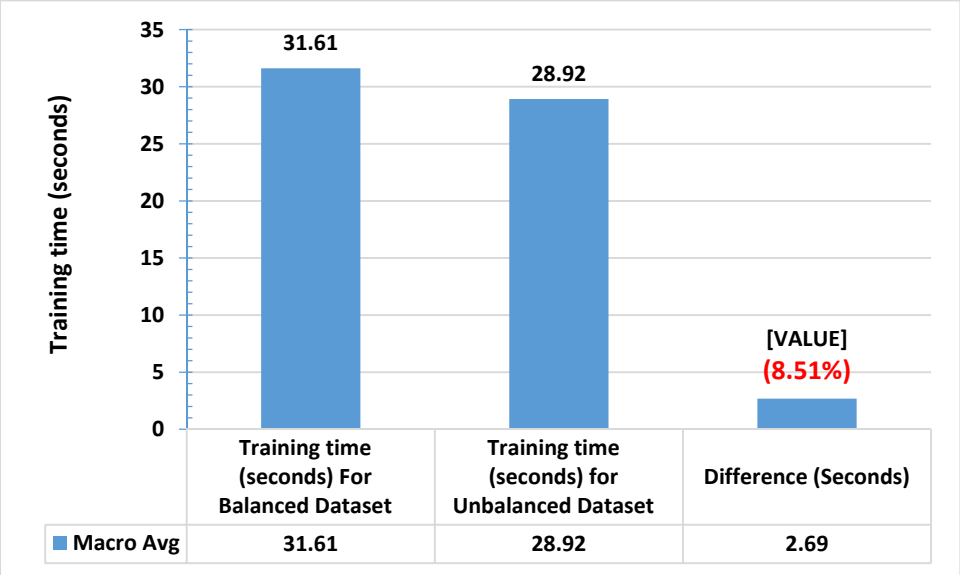

Figure 10 Comparison of the training time performance of the DTBC model for the balanced and the imbalanced dataset cases

4. Conclusion

The application of Decision Tree classifier for binary classification of the data packets in the Industrial Internet of Things (IIoT) network traffic is presented. The study focused on examining the impact of data balancing on the performance of the classifier model. The dataset was split and used to train the model without data balancing, afterwards, the data balancing was conducted and the model was again trained. The results showed that the attack class which is the minority class without the data balancing has significantly lower performance without the data balancing while the majority class, which is the normal data packet class has very slight reduction in performance in without data balancing, In all, the data balancing improved the overall accuracy by over 6 % while the execution time of the balanced dataset case is also higher by 8.51 %.

References

1. Rath, Kali Charan, Alex Khang, and Debanik Roy. "The role of Internet of Things (IoT) technology in Industry 4.0 economy." *Advanced IoT technologies and applications in the industry 4.0 digital economy*. CRC Press, 2024. 1-28.
2. Ahmed, Shams Forruque, et al. "Industrial Internet of Things enabled technologies, challenges, and future directions." *Computers and Electrical Engineering* 110 (2023): 108847.

3. Wójcicki, Krzysztof, et al. "Internet of things in industry: research profiling, application, challenges and opportunities—a review." *Energies* 15.5 (2022): 1806.
4. Peter, Onu, Anup Pradhan, and Charles Mbohwa. "Industrial internet of things (IIoT): opportunities, challenges, and requirements in manufacturing businesses in emerging economies." *Procedia Computer Science* 217 (2023): 856-865.
5. Ahmed, Shams Forruque, et al. "Industrial Internet of Things enabled technologies, challenges, and future directions." *Computers and Electrical Engineering* 110 (2023): 108847.
6. Chi, Hao Ran, et al. "A survey of network automation for industrial internet-of-things toward industry 5.0." *IEEE Transactions on Industrial Informatics* 19.2 (2022): 2065-2077.
7. Babayigit, Bilal, and Mohammed Abubaker. "Industrial internet of things: A review of improvements over traditional scada systems for industrial automation." *IEEE Systems Journal* 18.1 (2023): 120-133.
8. Demertzi, Vasiliki, Stavros Demertzis, and Konstantinos Demertzis. "An overview of privacy dimensions on the industrial internet of things (iiot)." *Algorithms* 16.8 (2023): 378.
9. Sarjan, Hamed, Amir Ameli, and Mohsen Ghafouri. "Cyber-security of industrial internet of things in electric power systems." *IEEE Access* 10 (2022): 92390-92409.
10. Xu, Jianpeng, et al. "Deep reinforcement learning for RIS-aided secure mobile edge computing in industrial Internet of Things." *IEEE Transactions on Industrial Informatics* 20.2 (2023): 2455-2464.
11. Aslam, Sidra, Mohammad MR Alshoweky, and Mohamed Saad. "Binary and multiclass classification of attacks in edge iiot networks." *2024 Advances in Science and Engineering Technology International Conferences (ASET)*. IEEE, 2024.
12. Ikram, Sumaiya Thaseen, et al. "Prediction of IIoT traffic using a modified whale optimization approach integrated with random forest classifier." *The Journal of Supercomputing* 78.8 (2022): 10725-10756.
13. Kumar, Vinod, et al. "Addressing binary classification over class imbalanced clinical datasets using computationally intelligent techniques." *Healthcare*. Vol. 10. No. 7. MDPI, 2022.
14. Dogra, Varun, et al. "A comparative analysis of machine learning models for banking news extraction by multiclass classification with imbalanced datasets of financial news: Challenges and solutions." *International Journal of Interactive Multimedia and Artificial Intelligence* 7.3 (2022): 35-52.
15. Thakkar, Ankit, and Ritika Lohiya. "Attack classification of imbalanced intrusion data for IoT network using ensemble-learning-based deep neural network." *IEEE Internet of Things Journal* 10.13 (2023): 11888-11895.

16. Hassan, Heba A., et al. "Detection of attacks on software defined networks using machine learning techniques and imbalanced data handling methods." *Security and Privacy* 7.2 (2024): e350.
17. Adi Pratama, Firza Refo, and Siskarossa Ika Oktora. "Synthetic Minority Over-sampling Technique (SMOTE) for handling imbalanced data in poverty classification." *Statistical Journal of the IAOS* 39.1 (2023): 233-239.