Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 13 Issue 3, March - 2026

Industrial Internet Of Things (IIoT) Networks Anomalous Packet Classification Based On Multi-Class Random Forest Model

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

Abstract

In this work, Industrial Internet of Things (IIoT) networks anomalous packet classification based on Multi-Class Random Forest (McRF) model is presented. The model helps to facilitate automated identification and classification of attacks on Industrial Internet of Things (IIoT) networks which is essential for robust intrusion detection and control on such network. The version of the case study X-IIoTID (eXtended Industrial Internet of Things Intrusion Dataset) used had Normal packet class with 421,417 instances which is about 51.3% of the entire dataset while the Attack packet class has 399,417 instances which is about 48.7% of the entire dataset. In view of the highly imbalanced data distribution among the attack classes and the normal class, the SMOTE method was used to balance the data sample instances for the minority classes. The results showed that the normal class had precision of 99.87 % while the RDOS, Exploitation and Exfiltration attack classes recorded precision of 100 %. Also, the normal class had Recall of 99.75 % while the C & C, RDOS, Weaponization and Exfiltration attack classes recorded Recall of 100 %. The model training time was 381.35 seconds, while accuracy was 99.78 % . Also, the overall precision of the McRF model was 99.60%, the Recall was 99.58%, and the F1 score was 99.59 % while the ROC UAC was 100 %. In all, the McRF model effectively identified the 9 different attack classes present in the IIoT traffic. The model is therefore recommended for use in intrusion detection framework for the IIoT network.

Keyword: Industrial Internet of Things (IIoT) Networks, Anomalous Packet Classification, Multi-Class Random Forest Model, X-IIoTID (eXtended Industrial Internet of Things Intrusion Dataset), Synthetic Minority Over-sampling Technique (SMOTE)

1. Introduction

Today, there is extensive investment in smart system in the industrial sector. This is changing the landscape in the industrial automation [1,2,3]. Robots, sensors and internet technologies are integrated to evolve highly efficient but complex industrial infrastructure that rely heavily on the secure communication among the diverse devices and systems that make up the industrial setup, otherwise known as Industrial Internet of Things (IIoT) [4,5,6,7]. To this end, any infraction on the data or communication among the diverse devices and systems in the industrial IIoT will result in extensive

damage and huge cost to the system [8,9,10]. As such, great effort is always made to mitigate any intrusion to the IIoT network. Such, intrusion detection setup requires threat identification and possibly classification unit which helps to detect and hence restrict the attack packets from gaining access to the IIoT network [11,12,13].

Accordingly, this work employed a Multi-class Random Forest (McRF) model [14,15] to classify the IIoT traffic into their respective attack classes, including the normal data class. The difficulty of data imbalanced is addressed using the Synthetic Minority Over-sampling Technique (SMOTE) [16]. The performance of the model is assessed in respect of Precision, Recall, and F1 Score, as well as Execution Time and Accuracy. The model is expected to serve in the intrusion detection module for IIoT networks.

2. Methodology

2.1 The multi-class Random Forest (McRF) model

In this work, multi-class Random Forest (McRF) model is used to identify and classify anomalous packets in an Industrial Internet of Things (IIoT) traffic into the various threat class categories. The the McRF model is a Random Forest Model (RFM) which is designed to classify data into more than two categories. The RFM architecture is presented in Figure 1. A typical RFM is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (classification) of the individual trees. This approach significantly reduces overfitting and improves robustness.

The RFM builds multiple decision trees, each trained on a random subset of the data and features. This ensemble approach reduces overfitting and improves performance by averaging the predictions from many trees. In this research, each individual tree uses a random selection of features from the total 79, promoting diversity among the trees. Each tree makes a prediction, and the most common outcome among all trees becomes the final prediction. The key parameter settings used for the DTM are as follows;

(i) random_state=42

(ii) n_estimators=100 → builds 100 trees

(iii) criterion='gini'

(iv) max_features='sqrt' → limits features considered at each split

(v) max_depth=None, min_samples_split=2, min_samples_leaf=1

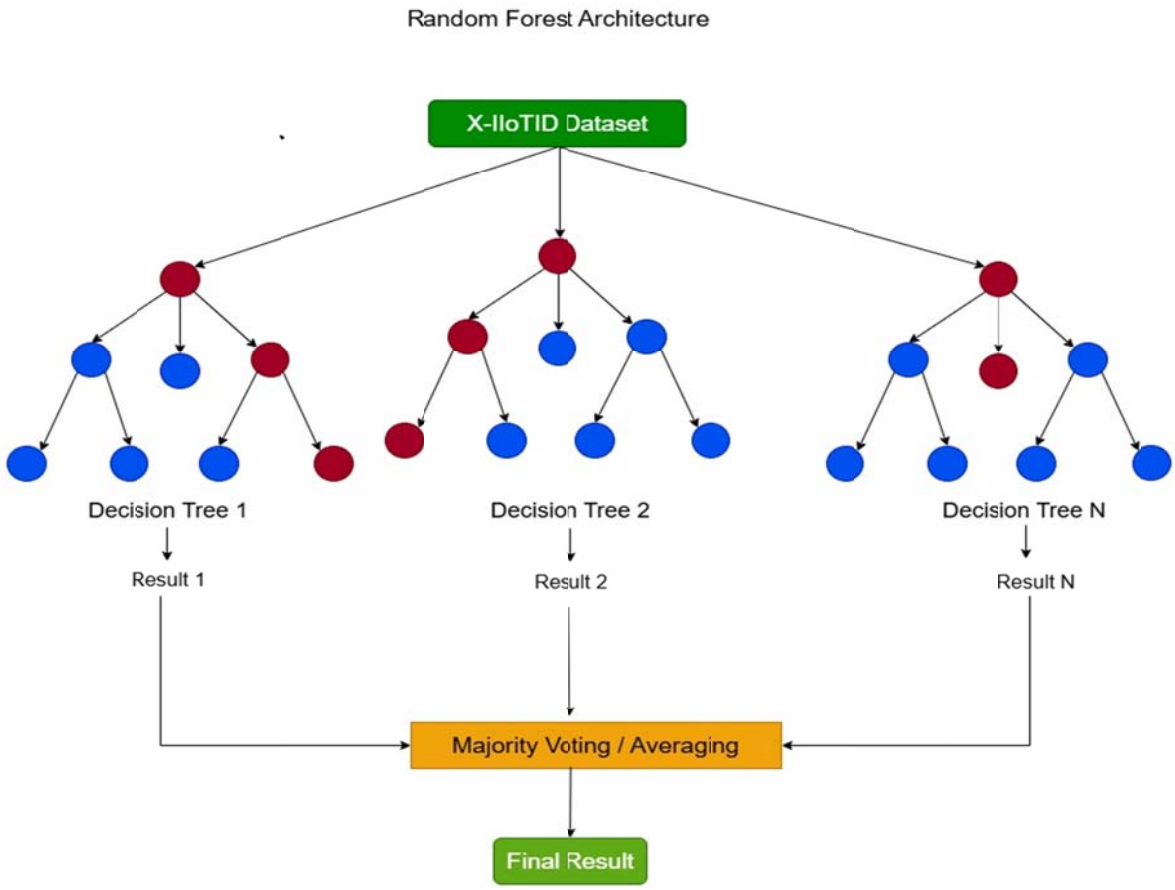

Figure 1 The Architecture of the Decision Tree Model

2.2 The case Study Dataset

The X-IIoTID (eXtended Industrial Internet of Things Intrusion Dataset) is used in the study. The version of X-IIoTID used consists of Normal packet with 421,417 instances which is about 51.3% of the entire dataset while the Attack has 399,417 instances which is about 48.7% of the entire dataset. However, the Attack class has several subclasses. In this study, 8 different attack classes and the normal class are considered, making a total of 9 classes for the multi-class McRF model. The class distribution of the data samples for the 9 classes in the dataset is shown in Table 1 and percentages of the class

distribution are shown in Figure 1. The Normal traffic is class 4 with the highest data sample of 243,220which is about 56 % of the whole dataset (Figure 1).

Table 1 The class distribution for the original dataset

Attack Class	Number of Samples before data balancing
Class 2	649
Class 0	1,551
Class 7	2,616
Class 1	11,385
Class 3	17,825
Class 8	26,685
Class 5	53,864
Class 6	65,762
Class 4	234,220

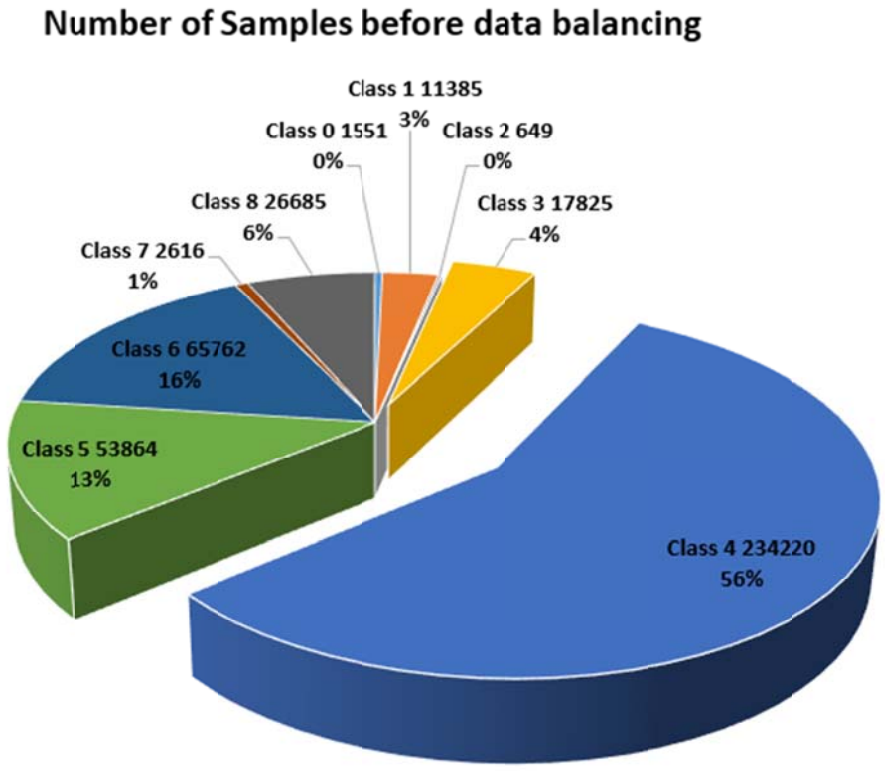

Figure 1 The distribution of the normal and attack samples in the dataset for the multi-class classification

2.3 Data Balancing using Synthetic Minority Over-sampling Technique (SMOTE)

In view of the highly imbalanced data distribution among the attack classes and the normal class, the Synthetic Minority Over-sampling Technique (SMOTE) approach is used to oversample the minority classes to equal the normal class in data sample size, as shown in Tale 2.

The summary of the multi-class classification resampling (class balancing using SMOTE), as captured in Table 2 is as follows:

(i) Original y_train_multiclass distribution: Counter({4: 234220, 6: 65762, 5: 53864, 8: 26685, 3: 17825, 1: 11385, 7: 2616, 0: 1551, 2: 649})

(ii) Resampled y_train_multiclass distribution: Counter({4: 234220, 8: 234220, 6: 234220, 5: 234220, 1: 234220, 3: 234220, 7: 234220, 0: 234220, 2: 234220}) (All classes balanced to 234220 samples each)

(iii) X_train_multiclass_resampled shape: (2107980, 79).

Tale 2 The summary of the multi-class classification resampling (class balancing using SMOTE)

Attack Class	Number of Samples before data balancing	Number of Samples after data balancing
Class 0	1551	234220
Class 1	11385	234220
Class 2	649	234220
Class 3	17825	234220
Class 4	234220	234220
Class 5	53864	234220
Class 6	65762	234220
Class 7	2616	234220
Class 8	26685	234220

2.4 Data Splitting for the Multi-class Classification Split (9 Granular Classes)

The dataset was split in 70/30 ratio. The summary of the multi-class dataset splitting is as follows:

(i) X_train_multiclass shape: (414557, 79)

(ii) X_ validation _multiclass shape: (177668, 79)

(iii) y_train_multiclass class distribution: Counter({4: 234220, 6: 65762, 5: 53864, 8: 26685, 3: 17825, 1: 11385, 7: 2616, 0: 1551, 2: 649})

(iv) y_validation_multiclass class distribution: Counter({4: 100380, 6: 28184, 5: 23085, 8: 11436, 3: 7640, 1: 4879, 7: 1121, 0: 665, 2: 278})

After the data splitting, the multi-class McRF model was trained with the training set and validated using the validation dataset. The McRF model is able to identify the 9 different attack classes and the performance of the model for each class is recorded. The model performance are captured in terms of Precision, Recall, and F1 Score, as well as Execution Time and Accuracy.

3. Results and discussion

The McRF model classification performance are presented in Figure 2, Figure 3, Figure 4, Figure 5, Figure 6, and Figure 7. The normal class has precision of 99.87 % while the RDOS, Exploitation and Exfiltration attack classes recorded precision of 100 % (Figure 2). The Lateral Movement Class recorded the least Precision score of 98.38%.

In Figure 3, the normal class has Recall of 99.75 % while the C & C, RDOS, Weaponization and Exfiltration attack classes recorded Recall of 100 % (Figure 3). The Exploitation Class recorded the least Recall score of 97.86%. Again, in Figure 4, the normal class has F1 Score of 99.81 % while the RDOS and Exfiltration attack classes recorded F1 Score of 100 % (Figure 4). The Lateral Movement Class recorded the least F1 Score of 98.69%. Furthermore, in Figure 5, the normal class has the highest normalized support of 56.77 % while the Exploitation Class recorded the least normalized support of 0.018 %. However, the model could identify the RDOS more than the normal class even when the RDOS has normalized support of 12.83 %.

The model training time is 381.35 seconds, as shown in Figure 6. The model accuracy is 99.78 % (Figure 7). Also, Figure 7 showed that the overall precision of the McRF model is 99.60%, the Recall is

99.58%, and the F1 score is 99.59 % while the ROC UAC is 100 %. In all, the McRF model demonstrated robust capability to identify the specific attack class from the case study IIoTID dataset.

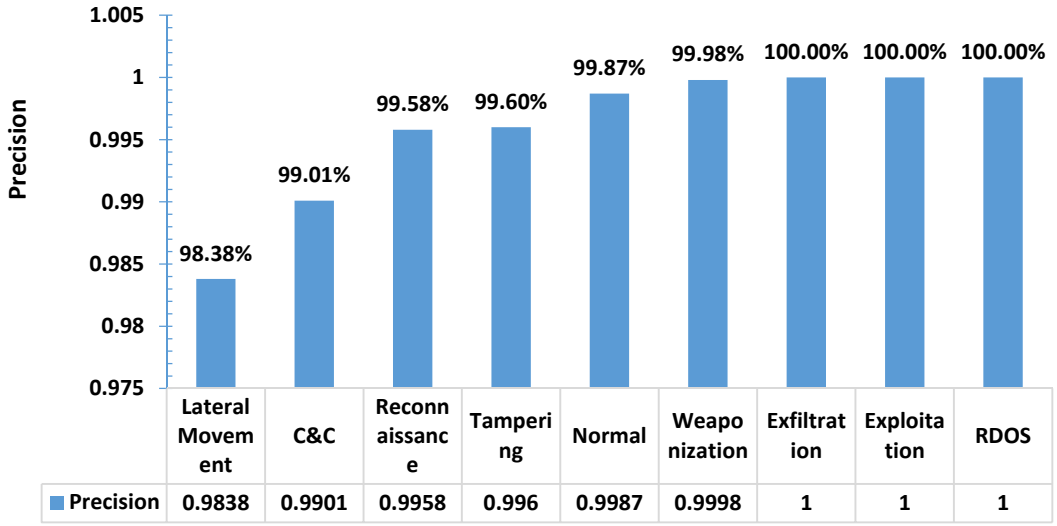

Figure 2 Comparison of the random forest precision results for the nine classes in the multiclass classification

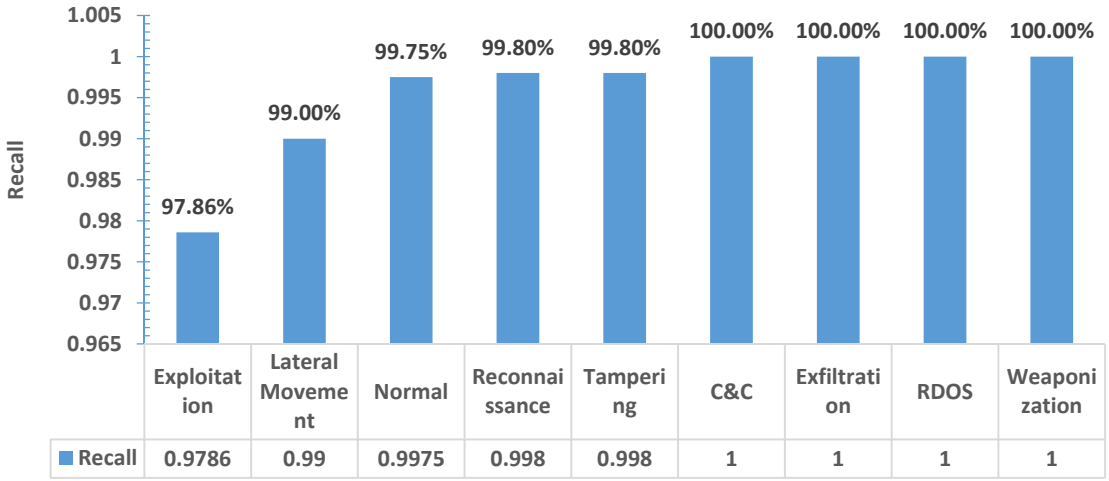

Figure 3 Comparison of the random forest recall results for the nine classes in the multiclass classification

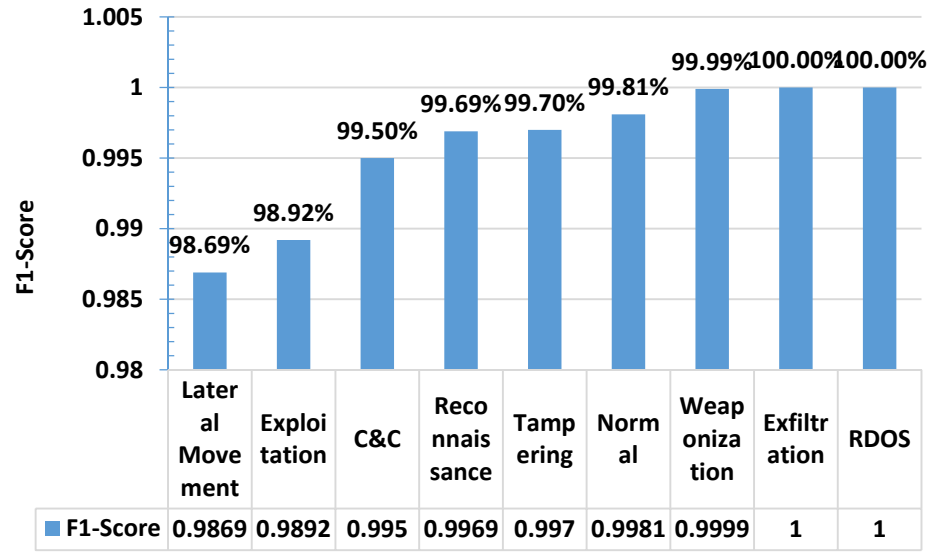

Figure 4 Comparison of the random forest F1-Score results for the nine classes in the multiclass classification

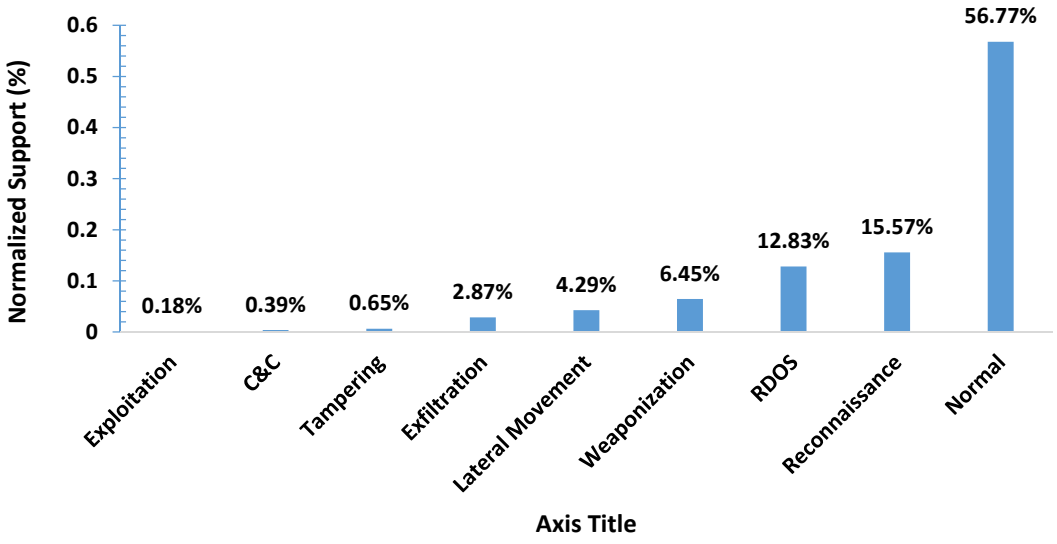

Figure 5 Comparison of the random forest normalized support results for the nine classes in the multiclass classification

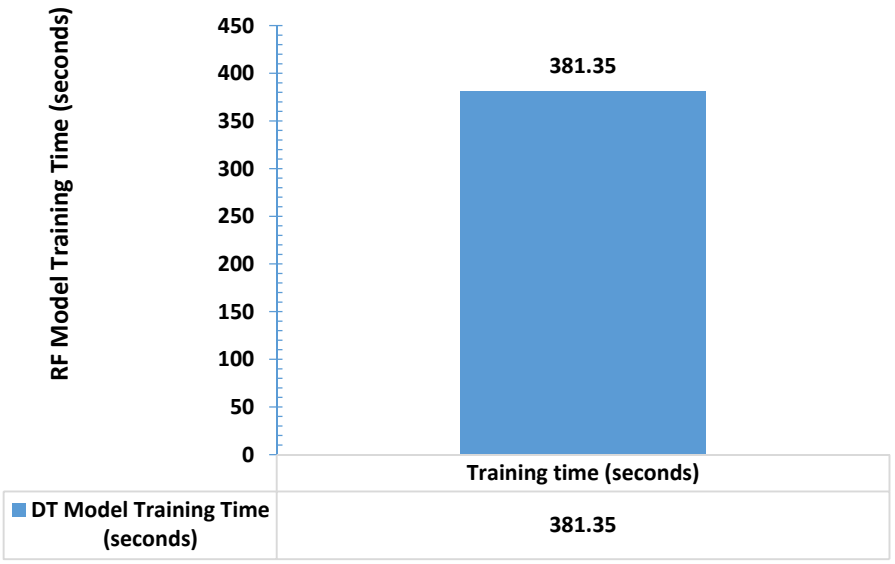

Figure 6 The random forest training time for the multiclass classification

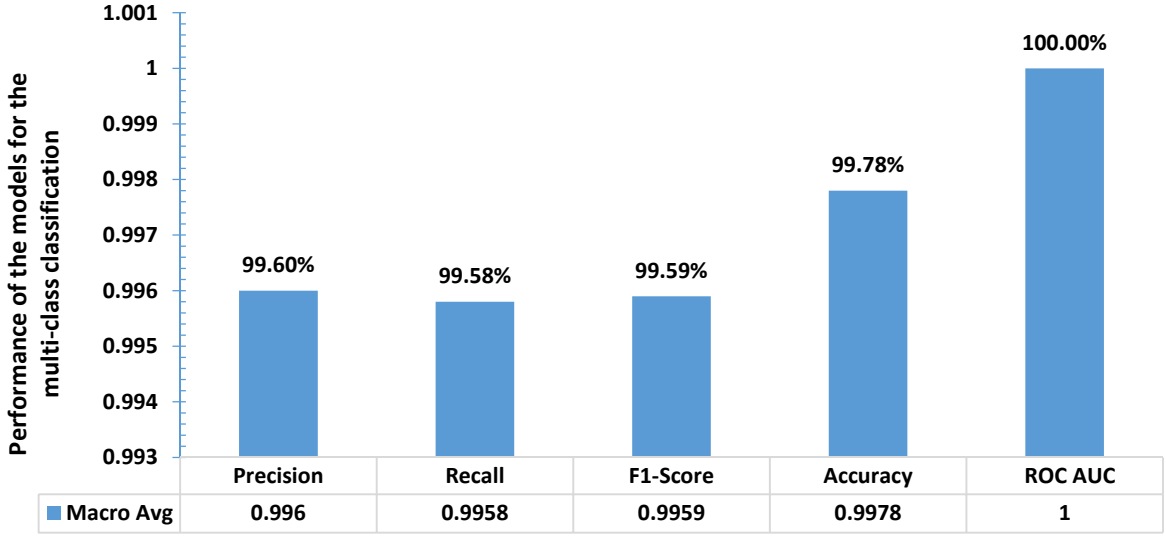

Figure 7 The mean values of the random forest performance parameters for the multiclass classification

4. Conclusion

The study considered Random Forest (RF) model for multi-class categorization of the attack classes in an X-IIoTID (eXtended Industrial Internet of Things Intrusion Dataset). The Multi-class Random Forest (McRF) model is trained and validated on the 70/30 split X-IIoTID dataset. The X-IIoTID dataset used had 9 different attack classes, including the normal class. The SMOTE (Synthetic Minority Over-sampling Technique) method was used to balance the data sample instances for the minority classes. The results showed that the accuracy of 99.78%.

References

1. Adel, Amr. "Unlocking the future: fostering human–machine collaboration and driving intelligent automation through industry 5.0 in smart cities." *Smart Cities* 6.5 (2023): 2742-2782.
2. Preuveneers, Davy, and Elisabeth Ilie-Zudor. "The intelligent industry of the future: A survey on emerging trends, research challenges and opportunities in Industry 4.0." *Journal of Ambient Intelligence and Smart Environments* 9.3 (2017): 287-298.
3. Sharma, Manjari, Tanmay Paliwal, and Payashwini Baniwal. "Challenges in Digital Transformation and Automation for Industry 4.0." *AI-Driven IoT Systems for Industry 4.0*. CRC Press, 2024. 143-163.
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. Kalsoom, Tahera, et al. "Impact of IoT on manufacturing industry 4.0: A new triangular systematic review." *Sustainability* 13.22 (2021): 12506.
6. Alabadi, Montdher, Adib Habbal, and Xian Wei. "Industrial internet of things: Requirements, architecture, challenges, and future research directions." *IEEE Access* 10 (2022): 66374-66400.
7. Behrendt, Andreas, et al. "Leveraging Industrial IoT and advanced technologies for digital transformation." *McKinsey & Company* (2021): 1-75.
8. Ahmed, Shams Forruque, et al. "Industrial Internet of Things enabled technologies, challenges, and future directions." *Computers and Electrical Engineering* 110 (2023): 108847.
9. Kebande, Victor R. "Industrial internet of things (IIoT) forensics: The forgotten concept in the race towards industry 4.0." *Forensic Science International: Reports* 5 (2022): 100257.
10. Khan, Izhar Ahmed, et al. "Enhancing IIoT networks protection: A robust security model for attack detection in Internet Industrial Control Systems." *Ad Hoc Networks* 134 (2022): 102930.
11. Alani, Mohammed M., and Ali Ismail Awad. "An intelligent two-layer intrusion detection system for the Internet of Things." *IEEE Transactions on Industrial Informatics* 19.1 (2022): 683-692.
12. Bansal, Komal, and Anita Singhrova. "Review on intrusion detection system for IoT/IIoT-brief study." *Multimedia Tools and Applications* 83.8 (2024): 23083-23108.
13. Alshahrani, Hani, et al. "Intrusion detection framework for industrial internet of things using software defined network." *Sustainability* 15.11 (2023): 9001.
14. Distefano, Veronica, Monica Palma, and Sandra De Iaco. "Multi-class random forest model to classify wastewater treatment imbalanced data." *Socio-Economic Planning Sciences* 95 (2024): 102021.
15. Farooq, Sheikh Afaan. "The multi-class detection of five stages of hepatitis c using the machine learning based random forest algorithm." *2023 world conference on communication & computing (WCONF)*. IEEE, 2023.
16. Amirruddin, Amiratul Diyana, et al. "Synthetic Minority Over-sampling TEchnique (SMOTE) and Logistic Model Tree (LMT)-Adaptive Boosting algorithms for classifying imbalanced datasets of nutrient and chlorophyll sufficiency levels of oil palm (Elaeis guineensis) using spectroradiometers and unmanned aerial vehicles." *Computers and Electronics in Agriculture* 193 (2022): 106646.