International Journal of Emerging Trends in Computer Science and Information Technology
ISSN: 3050-9246 | https://doi.org/10.63282/3050-9246.IJETCSIT-V2I2P110
Eureka Vision Publication | Volume 2, Issue 2, 83-91, 2021

Original Article

Anomaly Identification in IoT-Networks Using Artificial Intelligence-Based Data-Driven Techniques in Cloud Environmen

Dinesh Rajendran[1], Venkata Deepak Namburi[2], Aniruddha Arjun Singh Singh[3], Vetrivelan Tamilmani[4], Vaibhav Maniar[5], Rami Reddy Kothamaram[6]
[1]Coimbatore Institute of Technology, MSC. Software Engineering.
[2]University of Central Missouri, Department of Computer Science.
[3]ADP, Sr. Implementation Project Manager.
[4]Principal Consultant (SAP), Infosys Ltd.
[5]Oklahoma City University, MBA / Product Management.
[6]California University of management and science, MS in Computer Information systems.

Abstract - In the dynamic and ever-evolving realm of computer networks, cloud computing has emerged as a major player. Cloud networks provide on-demand access to pooled resources, but anomalies can compromise their security and integrity. Recognizing anomalies in network traffic is crucial for data security in modern networked systems including the Internet of Things (IoT). This research introduces a novel anomaly detection technique that utilizes the CICIDS-2017 dataset, the Random Forest (RF) classifier, and machine learning (ML). Using the Synthetic Minority Over-Sampling Technique (SMOTE), class balancing is accomplished. Some of the data pre-processing procedures included in the suggested methodology are one-hot encoding, feature selection, handling missing data, and z-score normalization. Two datasets are created from the processed data: one for training and another for testing. Using the Random Forest model, both normal and harmful traffic patterns can be identified. The experimental findings demonstrate that the model achieves a high level of resilience and reliability in detecting anomalies. It obtains an AUC of 1.00, an F1-score of 99.65%, a recall of 99.31%, a precision of 99.99%, and an accuracy of 99.65%. The proposed system offers adequate management of imbalanced and high-dimensional network data, making it applicable in real-world IoT settings. It provides scalability, flexibility, and high detection rates.

Keywords - Cybersecurity, Cloud, IOT, Anomaly Intrusion Detection, Machine Learning, Deep learning.

1. Introduction

Cloud computing is defined by its service-oriented delivery models and has become the standard for distributed, scalable, and on-demand computer services. Public, private, hybrid, and multi-cloud configurations all incorporate PaaS, IaaS, and SaaS, or software as a service, into their respective cloud computing architectures [1]. The inherent elasticity, resource abstraction, and multi-tenancy of cloud infrastructures have enabled unprecedented levels of agility and operational efficiency across various industry verticals. However, this architectural complexity introduces a broadened attack surface, where virtualized assets, containerized microservices, and software-defined networks (SDNs) collectively amplify the risk of cyber threats [2]. The stateless and dynamic nature of cloud-based workloads, combined with the heterogeneity of network endpoints and ephemeral resource provisioning presents unique challenges to conventional security mechanisms that rely on deterministic, perimeter-based protection models .

The Iao has completely transformed the way factories run by combining state-of-the-art IoT technologies with more conventional industrial automation systems. This convergence has enabled seamless connectivity between devices such as sensors, robots, mixing tanks, and control systems across sectors including energy, healthcare, and automotive [3]. The evolution of IIoT enables smart decision-making, real-time data collection, and advanced analytics, which in turn improve product quality, increase production efficiency, and decrease operational costs [4][5]. This research presents a smart Internet of Things (IoT) gadget that can autonomously detect anomalies on-site and is cost-effective, thus taking a step in that direction. Anomaly detection is a branch of artificial intelligence that compares new data sets or streams of data with previously collected ones to identify out-of-the-ordinary data patterns. In most cases, a model outlining the typical patterns of behavior is established before any outliers are detected. This allows for the possibility of seeing out-of-the-ordinary patterns, or those that deviate significantly from the "normal" model [6].

Intrusion detection systems (IDS) are only one of several recent uses where breakthroughs in machine learning have demonstrated efficacy. Security applications could benefit from learning-based approaches, as these models could be trained to handle complex and ever-changing data utilizing extensive datasets. Incorporating firewalls with these learning models can

enhance their efficiency. A well-trained model that includes all possible sorts of attacks greatly enhances the efficiency of anomaly detection while being reasonably complex and cost-effective.

1.1. Motivation and Contribution of Study

The growing popularity of networked systems like the IoT has led to a dramatic increase in network traffic, making these environments vulnerable to a wide range of cyberattacks. Conventional IDS often overreacts to skewed, high-dimensional data, missing threats, while producing more false positives. As a result, a trustworthy anomaly detection system that can handle complicated and ever-changing traffic patterns while swiftly identifying malicious actions in real-time is vital. This study is motivated by the need to improve detection reliability and performance using advanced machine learning techniques that can generalize well to diverse network scenarios. The following research contributions of this work are:

- Using the CICIDS-2017 dataset as a baseline for anomaly detection in IoT and network contexts, which includes actual benign and attack traffic patterns.
- Feature selection, missing value handling, z-score normalization, and SMOTE-based data balancing are some of the data preparation procedures used to address class imbalance and excessive dimensionality.
- Development of a Random Forest-based classification model, leveraging ensemble learning to improve prediction robustness, minimize overfitting, and ensure reliable detection of anomalous traffic.
- Several performance indicators, such as accuracy, precision, recall, F1-score, and ROC-AUC, are utilized by the framework to evaluate the model's effectiveness.
- The proposed design is beneficial as shown by results from improved performance metrics such as accuracy, precision, recall, F1-score, and AUC.

1.2. Novelty and justification

Anomaly detection systems are designed to meet the growing need for robust security measures in networked and Internet of Things (IoT) environments, where sophisticated attacks pose a threat to sensitive data and operations. Issues including class imbalance, high-dimensional data, and poor interpretability are common with current methods. This study integrates a Random Forest classifier with comprehensive pre-processing, feature engineering, z-score normalization, and SMOTE-based data balancing to achieve accurate and unbiased detection of anomalous traffic. The novelty of the work lies in the combination of an ensemble learning model with systematic data preparation techniques, resulting in near-perfect accuracy and AUC on the CICIDS-2017 dataset. Furthermore, the framework is scalable, adaptable to dynamic IoT traffic patterns, and provides a reliable solution for real-time anomaly detection, distinguishing it from conventional ML and DL methods.

1.3. Structure of Paper

The paper's overall outline is as follows In Section II, survey the research on cloud computing outliers. In Section III, go over the proposed methodology. In Section IV, provide the results and comparisons, and in Section V, draw some conclusions and offer some recommendations for further research.

2. Literature Review

The cloud anomaly detection system is the subject of this section's research. Makes use of a variety of machine learning methods; Table I provides a review of this research.

Wani et al. (2019) Distributed denial of service attacks are critical assaults that put network availability at risk. Identifying and responding to cyberattacks has become increasingly challenging due to their growing sophistication and frequency. An attack tool called Tor Hammer was used to investigate the own cloud environment, and a new dataset was constructed using Intrusion Detection System. Several ML algorithms are employed in this context SVM achieved a classification accuracy of 99.7%, NB of 97.5%, and RF of 98.0% [7].

Garg et al. (2019) suggest a model designed to efficiently detect anomalies in networks, which operates in two stages. In the first step, feature selection using ImGWO is done to find the best compromise between two goals: minimizing the set of features and reducing the error rate. The second phase is classifying network anomalies using ImCNN. The effectiveness of the suggested methodology is demonstrated using synthetic and benchmark datasets such as DARPA'98 and KDD'99. In terms of F-score, false positive rate, detection rate, and accuracy, the results demonstrate that the proposed cloud-based anomaly detection model surpasses the state-of-the-art models used for network anomaly detection. Improved accuracy (3.62%), false positive rate (4.08%), and detection rate (8.25%) are some ways in which the proposed model surpasses traditional GWO with CNN [8].

Kotenko, Saenko, and Ageev (2018) This involves the development and evaluation of an innovative algorithm for analyzing network data in real-time or near-real-time. Also cover a variety of uses for intelligent agents that monitor data transfer speeds on IoT networks Integrated circuits (ICs), embedded devices (EDs), and supercomputers (SCs) come in that order. The agents are constructed using the pseudo-gradient anomaly detection method and fuzzy logic inference. There is real-time processing capability in the suggested algorithm. A 90% speedup and a 50% improvement in accuracy are revealed by the experimental examination of the procedure [9].

Dutt, Borah and Maitra (2018) The dangers can be categorized into patterns or they can be completely random. This means that the data coming into the network is quite diverse and packed with features. Before using the filtered collection of patterns to detect unknown threats, the first step is to minimize the number of patterns. Using machine learning methods and Principal Component Analysis (PCA) in a WEKA environment, this research presents a strategy to construct an Intrusion Detection System (IDS). The method's enhanced detection effectiveness makes it more effective. Compared to the current methods, the results show an increase in true positives and a decrease in false positives [10].

Salman et al. (2017) Research material today tends to focus on finding anomalies and putting them into groups, rather than just finding them. Using widely-used, publicly-available data, have built and tested learning models for attack detection and classification. The goal of using LR and RF, two supervised machine learning approaches, was to achieve precision. Because assaults are similar, it is still feasible to achieve less accurate classification even with complete detection. The accuracy of categorization is 93.6% and detection accuracy is greater than 99% according to results [11].

Mehmood and Rais (2016) use an SVM ant system, which is a special case of ant colony optimization, to find and remove unnecessary characteristics that improve the SVM classification algorithm's accuracy. Anomaly detection uses benchmark datasets, such as KDD99. Of the 41 features used to model all instances in KDD99, a few are superfluous or even useless. The ant system culled the design to get rid of all the extraneous parts. After selecting a subset of features using the ant system, support vector machine is used for further testing. Experiment results demonstrated that the classification system trained using the reduced feature set performed much better. The accuracy level, false positive rate, and true positive rate are the metrics used in this assessment [12].

Table 1. Summary of Related Works on Anomaly Detection in IOT

Author & Year	Methodology	Result	Key Findings	Limitation/Future Work
Wani et al. (2019)	Used Tor Hammer to generate dataset in ownCloud; applied ML algorithms (SVM, NB, RF) for classification.	Accuracy: SVM 99.7%, RF 97.6%, NB 98.0%	Showed high efficiency of SVM for DDoS detection compared to RF and NB.	Dataset limited to ownCloud environment; needs validation on larger and diverse datasets.
Garg et al. (2019)	Suggested two-stage model: ImGWO to select features and ImCNN to classify; tested on DARPA 98 and KDD 99, and synthetic data.	Avg. improvements: Detection rate +8.25%, False Positives −4.08%, Accuracy +3.62%.	Hybrid cloud-based anomaly detection model outperforms standard GWO+CNN.	Limited to benchmark datasets; future work can extend to real-time cloud traffic.
Kotenko, Saenko & Ageev (2018)	Developed pseudo-gradient anomaly detection with fuzzy logical inference for real-time IoT networks; tested on various devices.	Gain: +50% in accuracy, +90% in speed.	Intelligent agents enable real-time network analysis on IoT.	Needs deployment in large-scale IoT with heterogeneous traffic.
Dutt, Borah & Maitra (2018)	Used PCA + ML algorithms in WEKA to reduce dimensionality and detect unknown threats.	Increased true positives and decreased false positives.	PCA improves IDS efficiency by filtering patterns before classification.	Requires evaluation on dynamic datasets with evolving threats.
Salman et al. (2017)	RF and Applied LR on public dataset to detect anomalies and categorization.	Detection accuracy >99%, Categorization accuracy 93.6%.	Highlighted difficulty in categorizing similar attacks despite high detection.	Needs methods to improve categorization accuracy for similar attack types.
Mehmood & Rais (2016)	Proposed Ant System + SVM; ant system for feature selection on KDD99 dataset.	Improved SVM performance with reduced feature set.	Feature optimization via ant system enhances detection accuracy.	Limited to KDD99 dataset; needs testing on modern datasets.

3. Methodology

The anomaly detection system was developed using the CICIDS-2017 dataset, with the process structured into distinct phases. Data pre-processing included activities such as handling missing values, identifying features, encoding them simultaneously, and normalizing them using z-score standardization; this was the initial stage in ensuring the data's quality and consistency. Attempted to level the playing field by employing the SMOTE. Then, the pre-processed dataset was split in half, with 80% going to the training subset and 20% to the testing subset. An RF classifier was employed for the analysis, trained on 80% of the data and tested on the remaining 20%. Accuracy, Precision, Recall, and F1-score were some of the metrics used to

evaluate the model's performance in an evaluation matrix. Subsequently, the outcomes underwent a comprehensive examination. Figure 1 illustrates the flowchart for anomaly detection.

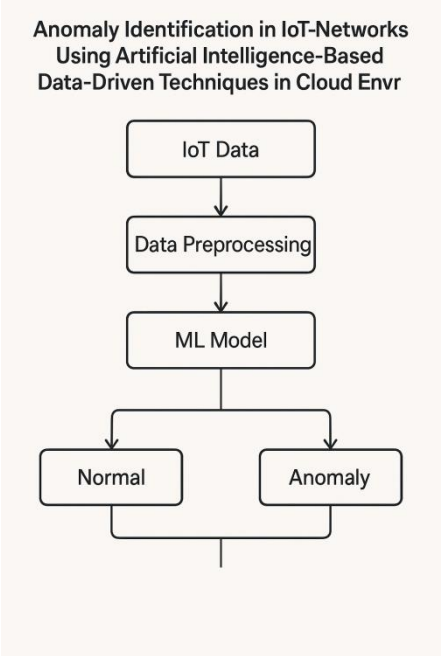

Figure 1. Flowchart of the Anomaly Detection System using ML Models

These methodology steps are discussed below in brief:

3.1. Data Collection and Visualization

ML to identify potential anomaly hazards relies on datasets. A dataset representing real-time network traffic that includes both benign and malicious behaviors, the used CICIDS-2017, is used in this study. Collecting background traffic in real-time is the primary focus of this dataset creation effort. The email, FTP, SSH, HTTP, and HTTPS protocols are used to identify 25 users in this harmless traffic. Over the course of five days, monitored the network traffic and inserted malicious traffic on some days while leaving the others alone. The intruder attacks that were experienced were Strikeforce FTP, Bruteforce SSH, 'DoS,' Heartbleed, Web Attack, Infiltration, Botnet, and DDoS. Take a look at these visual representations of the CICIDS-2017 data set.

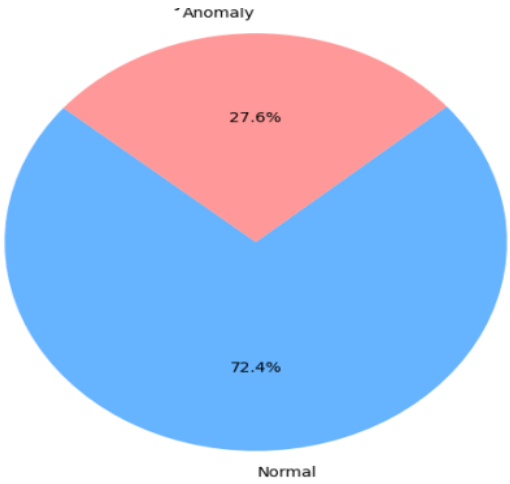

Figure 2. Anomaly Distribution Pie Chart

Figure 2 presents Anomaly Distribution Pie Chart. The chart is a circular pie graph divided into two sections. The larger section, colored light blue, represents the "Normal" class and accounts for 72.4% of the data. The smaller section, colored light red, represents the "Anomaly" class and makes up 27.6% of the data. A large portion of the data falls into the "Normal" category, as is visually shown, indicating that the dataset is imbalanced.

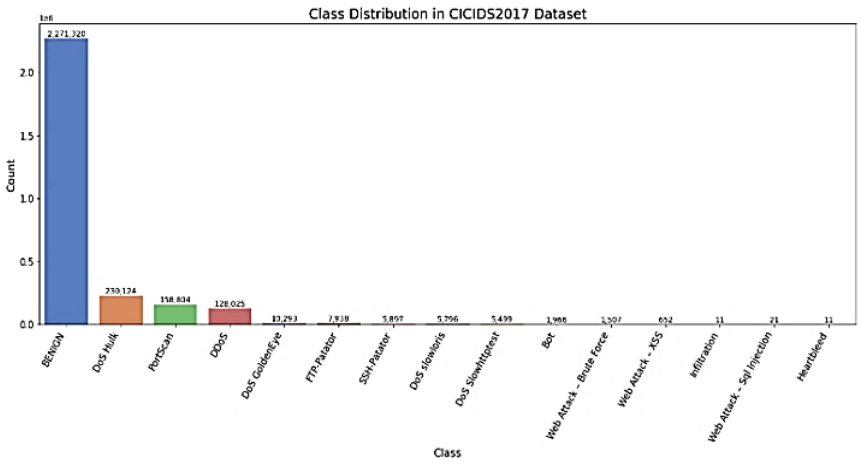

Figure 3. Class Distribution in CICIDS2017 Dataset

The initial distribution of classes in the CICIDS2017 dataset, as shown in Figure 3, reveals a striking difference between various classes. The benign class dominates with over 2.2 million instances, followed by DoS Hulk with around 230,124 and PortScan with approximately 158,804 samples. Some attack types, such Heartbleed, Botnet, and Infiltration, have as low as twenty samples. Anomaly detection that is both fair and effective requires data balancing strategies to prevent biased model training, in which the classifier favors majority classifications.

3.2. Data Pre-Processing

The dataset contained some missing values, which were addressed in the first step of data preprocessing. The CICIDS2017 dataset is expected to retain its general dependability and quality when these missing values are removed, thanks to its enormous volume. Used the data. isnull () .sum() function as a check to make sure the dataset was complete and free of missing entries. This processed the data set by dividing numerical values, normalizing the data, and balancing the data.

3.2.1. Feature Selection

Feature selection techniques are designed to boost classification performance by identifying and utilizing the most important characteristics. This allows ML-based cybersecurity systems to make more accurate predictions, as each feature contributes unique insights. There are two main categories of feature selection methods: filter methods that use mutual information and correlation, and wrapper methods that use recursive feature elimination.

3.2.2. One Hot Encoding

Data preprocessing techniques like one-hot encoding are useful for making numerical formats out of category variables, which are more amenable to machine learning algorithms. The method is effective because it uses one-hot encoding to represent categorical data, which is done by establishing binary (0 or 1) columns for every distinct category in a feature.

3.2.3. Normalization with Z-score

Normalizing the numerically processed data can prevent gradient dispersion, which is caused by the significant variation in individual features, when employing the backpropagation algorithm. All of the CICIDS2017 data is transformed using the z-score normalization approach. Equation (1) is used to derive it.

$$m'i = \frac{m_i - \overline{m}}{x}$$

The values of the data sample before and after normalization are represented by m_i and $m'i$, respectively. A feature's average data value before normalization is denoted by $\overline{m}$.

3.3. Data balancing with SMOTE

The SMOTE method is a type of oversampling that is often used in medical applications to fix class-imbalanced data. By creating synthetic data points from its closest neighbours using Euclidean distance, a huge number of data instances in the minority class are increased. This process is outlined in the paper. Since these new instances are created using the original features, they are certain to look just like the original data. SMOTE can introduce extra noise, making it a less optimal choice for high-dimensional data. This research utilizes the SMOTE method to create a brand-new training dataset.

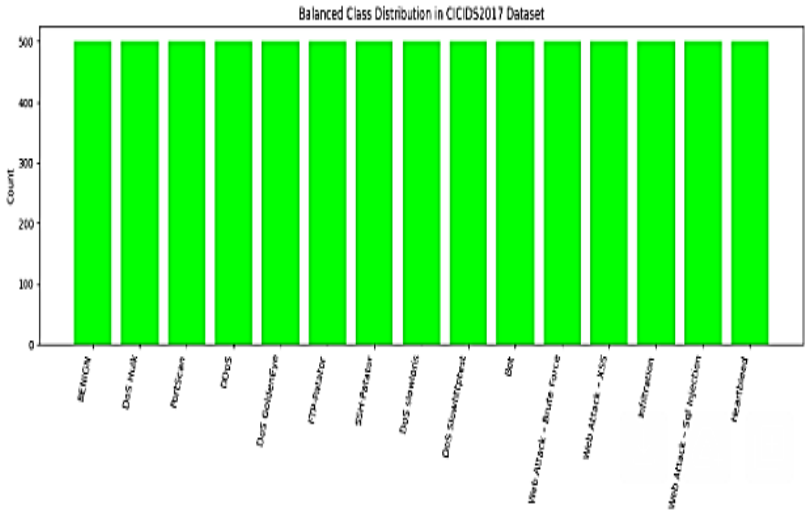

Figure 4. After Data Balanced Graph in CICIDS2017 Dataset

The class distribution of the CICIDS2017 dataset after using balancing strategies is shown in Figure 4. Critical for enhancing the efficiency of machine learning models, this balanced distribution guarantees that the dataset is free from class imbalance problems. A balanced dataset improves the reliability of anomaly detection in network traffic, makes training and evaluation more fair, and lowers model bias. It does this by providing equal representation of different attack types alongside the benign class, including DoS Hulk, Ports can, Botnet, Web Attacks, Infiltration, and others.

3.4. Data Splitting

Training and testing subsets were created from the dataset to enable effective training and evaluation of models. In this study, training was given to 80% of the data and testing was given to the other 20%.

3.5. Classification with Random Forest Model

Decision Trees are binary classifiers that may be found in Random Forests. Every decision tree is structured with a root node, numerous internal split nodes, and leaf nodes that are used to classify events [13]. Random Forest's classification is up for grabs in every decision tree's vote. Since RFs converge quickly and are resistant to over-fitting, they can be useful in many applications. In addition to data classification, RF provides a way to quantify the importance of features in making the classification call. This is achieved by calculating the Gini index and the accuracy loss when a feature is not used for classification, respectively. The Gini index quantifies how homogeneous a set of data is when partitioned by a specific characteristic. When the Gini index drops significantly, it indicates that the attribute is highly important for accurate categorization. Equation (2) represents the Random Forest's classification function:

$$\hat{y} - mode\{h_t(x), t = 1,2, \dots T\}$$

Where $\hat{y}$ is the final projected class, T is the total number of trees in the forest, and $h_t(x)$is the forecast of the tth decision tree.

3.6. Performance Matrix

Using evaluation metrics, anomaly detection algorithms can be efficiently tested. These parameters affect the model's outlier detection accuracy. Criteria such as accuracy, precision, recall, and F1 score are crucial. Precisely speaking, the accuracy rate is the proportion of valid outliers to the overall number of outliers. The term "recall" describes the percentage of anomalies that were really found. The F1 score finds a happy medium between recall and precision with its harmonic mean. A well-classified dataset is one with a high degree of accuracy [14]. The mathematical formulae for these evaluation metrics are Equation. (3) (5):

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1 - score = 2 * \frac{(precision*recall}{(precision+recall)} \quad (6)$$

The number of true negatives, false positives, and true positives relative to each other are represented by the variables FP, FP, and TN, respectively.

The AUC is a ROC plot that considers both the true positive rate (recall) and the FPR to determine a classifier's accuracy. A model's overall performance can be measured by its AUC:

$$AUC = \int_0^1 TPR(x)dx$$

A bigger area under the curve (AUC), as indicated in Equation. (7), indicates that a model can differentiate between classes.

4. Results and Discussion

Results and analysis related to the model's execution are provided in this section. To do this, utilized the Python programming language and the Jupyter Notebook software. Data pre-processing and model implementation were handled by the scikit-learn and Keras packages, respectively. Used a MacBook Air equipped with 8 GB of RAM and an Intel Core i3 CPU 1.6 GHz to train the suggested model. Table II shows the outcomes of the CICIDS2017 dataset's RF model run by an anomaly detection system. A 99.65% F1-score, 99.65% accuracy, 99.99% precision, and 99.31% recall show that the RF model performs quite well. Additionally, the model's optimal ROC of 1.00 further supports its extraordinary capability to discern between typical and unexpected traffic. As a result of these results, the RF classifier is a strong contender for intrusion detection systems since it reliably and effectively detects abnormalities in network traffic.

Table 2. Random Forest Model performance for anomaly detection system on CICIDS2017dataset

Performance	RF
Accuracy	99.65
Precision	99.99
Recall	99.31
F1 score	99.65
ROC	1.00

RF:	precision	recall	f1-score	support
0	0.99	1.00	1.00	724
1	1.00	0.99	1.00	724
accuracy			1.00	1448
macro avg	1.00	1.00	1.00	1448
weighted avg	1.00	1.00	1.00	1448

Figure 5. Classification Report of RF Model

Figure 5 displays the results of the RF model's categorization efforts; with an overall accuracy of 1.00, all 1448 samples were correctly identified. With a recall of 1.00 and a precision of 0.99, classes 0 and 1 respectively, exhibit strong F1-scores. Both the macro and weighted average scores of 1.00 in all metrics, as well as the balanced distribution of samples (724 in each class), lend credence to this outstanding accomplishment.

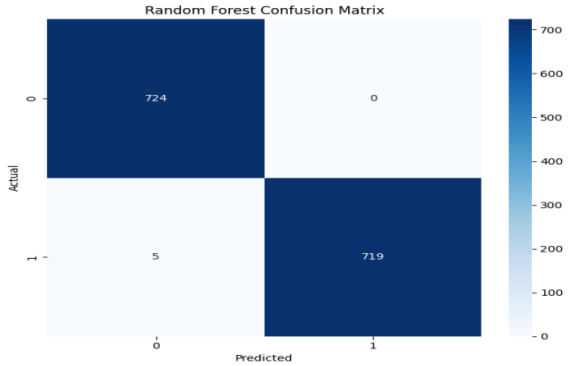

Figure 6. Confusion Matrix of RF Model

Figure 6 RF model's confusion matrix reveals the accuracy of the model's classification: 724 out of 1448 samples were assigned to the correct class 0 and 719 to the correct class 1, according to the model. According to the template, the model had 5 false negatives (class 1 samples wrongly forecasted as class 0), but no false positives (class 0 samples improperly predicted as

class 1) overall. The total result of the matrix suggests a highly precise model that successfully categorized the overwhelming majority of the samples, proving its high predictive power with a very low number of failures.

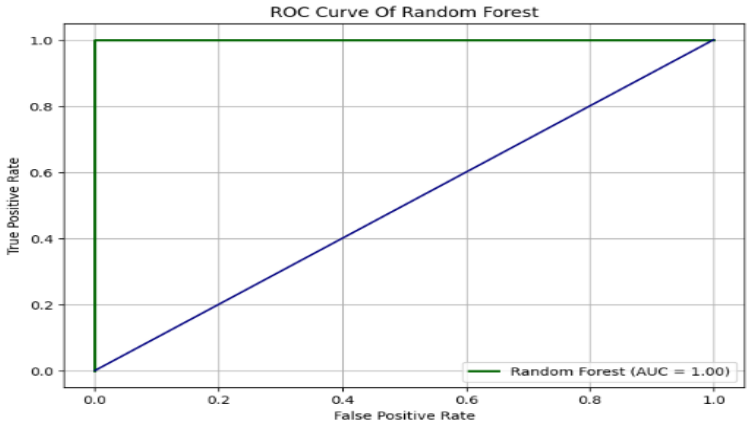

Figure 7. ROC Curve Graph of Random Forest Model

Figure 7 displays the ROC curve of the RF model, which illustrates its performance across different levels of classification To depict a model that successfully differentiates between the two classes, an ideal curve borders the top-left side of the graph that compares the True Positive rate with the False Positive rate. This is also backed by the value of the AUC being 1.00, which affirms the fact that the model is fully competent to give a distinct separation between the positive and negative classes and has the ideal performance based on classification.

Table 3. Comparison between ML and DL models for anomaly detection in Internet of Things

Performance	**RF**	**CBF** [15]	**MLP**[16]
Accuracy	99.65	68.35	84
Precision	99.99	67.69	88
Recall	99.31	67.58	84
F1 score	99.65	67.308	83

Table III provides a comparative study of various ML and DL frameworks relating to anomaly detection in the IoT. In contrast to the Confidence-Based Filtering (CBF) method and MLP, the RF model significantly outperforms both with figures like 99.65% accuracy, 99.99% precision, 99.31% recall, and 99.65% F1-score. Compared to the CBF strategy, which performs the worst, MLP performs averagely with 84% accuracy, 88% precision, 84% recall, and 83% F1-score. In contrast, the CBF approach fares the worst with 67.30% F1-score, 67.58% recall, 67.69% precision, and 67.35% accuracy. Based on these results, RF is more effective than traditional and deep learning-based approaches in achieving solid and reliable anomaly detection in IoT environments.

The suggested model for IoT anomaly detection is built on a robust ML model that ensures high recall, accuracy, precision, and F1-score; this makes it more than enough to identify malicious or aberrant behavior in massive IoT networks. The model uses advanced detection based on the capacity to work with multi-dimensional data with high complexity and reduce false positives and negatives. Its main strength is its scalability, dynamism to changing patterns of IoT traffic, and capability to deliver real-time insights, all of which are essential to ensuring security and resilience in resource-limited IoT ecosystems.

5. Conclusion

Cybersecurity refers to the integration of many procedures, technologies, and practices that aim to safeguard information systems, networks, programs, and data from intrusion, attacks, or destruction. Constant accessibility, uncompromised data integrity, and privacy are its stipulations. In recent years, anomaly detection in cloud computing has emerged as a valuable application of ML, providing effective tools to counter evolving security challenges. This paper presents an anomaly detection system developed using the CICIDS-2017 dataset and a Random Forest classifier. Utilizing systematic preprocessing approaches such feature selection, one-hot encoding, z-score normalization, and SMOTE helped alleviate class imbalance and improve efficiency. Accuracy, precision, recall, F1-score, and AUC were all above average for the trained Random Forest model, which achieved 99.65%. This approach holds great promise for intrusion detection in the IoT and other areas of network security, as evidenced by these results, which demonstrate its ability to reliably distinguish between regular and aberrant data. For future work, advanced DL models such as LSTMs, CNNs, and attention-based architectures could be explored to capture temporal and spatial patterns in network traffic. Real-time deployment in IoT or edge environments, the integration of Explainable AI (XAI) for interpretability, and enhanced robustness against adversarial attacks remain critical avenues for further research.

References

[1] Gopi, "Zero Trust Security Architectures for Large-Scale Cloud Workloads," *Int. J. Res. Anal. Rev.*, vol. 5, no. 2, pp. 960–965, 2018.

[2] S. S. S. Neeli, "Serverless Databases: A Cost-Effective and Scalable Solution," *Int. J. Innov. Res. Eng. Multidiscip. Phys. Sci.*, vol. 7, no. 6, p. 7, 2019.

[3] H. Chen, M. Hu, H. Yan, and P. Yu, "Research on industrial internet of things security architecture and protection strategy," in *Proceedings - 2019 International Conference on Virtual Reality and Intelligent Systems, ICVRIS 2019*, 2019. doi: 10.1109/ICVRIS.2019.00095.

[4] Y. N. Soe, Y. Feng, P. I. Santosa, R. Hartanto, and K. Sakurai, "Rule Generation for Signature Based Detection Systems of Cyber Attacks in IoT Environments," *Bull. Networking, Comput. Syst. Softw.*, vol. 8, 2019.

[5] S. S. S. Neeli, "The Significance of NoSQL Databases : Strategic Business Approaches and Management Techniques," *J. Adv. Dev. Res.*, vol. 10, no. 1, p. 11, 2019.

[6] M. Antonini, M. Vecchio, F. Antonelli, P. Ducange, and C. Perera, "Smart Audio Sensors in the Internet of Things Edge for Anomaly Detection," *IEEE Access*, vol. 6, pp. 67594–67610, 2018, doi: 10.1109/ACCESS.2018.2877523.

[7] A. R. Wani, Q. P. Rana, U. Saxena, and N. Pandey, "Analysis and Detection of DDoS Attacks on Cloud Computing Environment using Machine Learning Techniques," in *Proceedings - 2019 Amity International Conference on Artificial Intelligence, AICAI 2019*, 2019. doi: 10.1109/AICAI.2019.8701238.

[8] S. Garg, K. Kaur, N. Kumar, G. Kaddoum, A. Y. Zomaya, and R. Ranjan, "A Hybrid Deep Learning-Based Model for Anomaly Detection in Cloud Datacenter Networks," *IEEE Trans. Netw. Serv. Manag.*, vol. 16, no. 3, pp. 924–935, 2019, doi: 10.1109/TNSM.2019.2927886.

[9] I. Kotenko, I. Saenko, and S. Ageev, "Applying intelligent agents for anomaly detection of network traffic in internet of things networks," in *Proceedings - 2018 IEEE International Conference on Internet of Things and Intelligence System, IOTAIS 2018*, 2018. doi: 10.1109/IOTAIS.2018.8600867.

[10] I. Dutt, S. Borah, and I. Maitra, "A Proposed Machine Learning based Scheme for Intrusion Detection," in *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, IEEE, Mar. 2018, pp. 479–483. doi: 10.1109/ICECA.2018.8474803.

[11] T. Salman, D. Bhamare, A. Erbad, R. Jain, and M. Samaka, "Machine Learning for Anomaly Detection and Categorization in Multi-Cloud Environments," in *2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud)*, IEEE, Jun. 2017, pp. 97–103. doi: 10.1109/CSCloud.2017.15.

[12] T. Mehmood and H. B. M. Rais, "SVM for network anomaly detection using ACO feature subset," in *2015 International Symposium on Mathematical Sciences and Computing Research, iSMSC 2015 - Proceedings*, 2016. doi: 10.1109/ISMSC.2015.7594039.

[13] S. D. D. Anton, S. Sinha, and H. Dieter Schotten, "Anomaly-based Intrusion Detection in Industrial Data with SVM and Random Forests," in *2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, IEEE, Sep. 2019, pp. 1–6. doi: 10.23919/SOFTCOM.2019.8903672.

[14] S. H. Kok, A. Abdullah, N. Z. Jhanjhi, and M. Supramaniam, "A review of intrusion detection system using machine learning approach," *Int. J. Eng. Res. Technol.*, vol. 12, no. 1, pp. 8–15, 2019.

[15] R. R. Palle, "Hybrid Multi-Objective Deep Learning Model for Anomaly Detection in Cloud Computing Environment," *Int. J. Sci. Res. Sci. Eng. Technol.*, vol. 1, no. 3, pp. 440–456, 2015.

[16] J. Alsamiri and A. Khalid, "Internet of things cyber attacks detection using machine learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 12, p. 627, 2019, doi: 10.14569/ijacsa.2019.0101280.