AnDet: ML-Based Anomaly Detection of UEs in a Multi-cell B5G Mobile Network for Improved QoS

Amar Sinha[1], Anmol Agrawal[2], Sandip Roy[3],
Venkanna Uduthalapally[1], Debanjan Das[2], Rajarshi Mahapatra[2], Sachin Shetty[3]
[1]*Department of Computer Science and Engineering*, IIIT Naya Raipur, India
[2]*Department of Electronics and Communication Engineering*, IIIT Naya Raipur, India
[3]*Center for Secure & Intelligent Critical Systems*, Old Dominion University, Virginia, USA
Email: amar@iiitnr.edu.in, anmol21101@iiitnr.edu.in, sroy@odu.edu,
venkannau@iiitnr.edu.in, debanjan@iiitnr.edu.in, rajarshi@iiitnr.edu.in, sshetty@odu.edu

***Abstract*—The rise of 5G technology has revolutionized wireless communication, ushering in a new era of unparalleled connectivity. Approaching Beyond 5G networks, Software Defined Networks have emerged as a promising paradigm to enhance flexibility and scalability in cellular networks. However, the challenge of 5G is to ensure the network's performance based on the different Quality of Service requirements for offering interactive services. Proactive detection of anomalies can significantly improve network performance. To address this issue, we present AnDet, a robust ML-based model specifically designed for anomaly detection in SDN-enabled B5G cellular networks. Next, we deploy the proposed AnDet for anomaly detection on SDN-enabled B5G cellular networks, enhancing overall network performance and reliability. Moreover, we incorporate Explainable AI approaches into the AnDet model to ensure transparency and interpretability. Finally, we capture different performance metrics for models and deployments to evaluate the proposed solution. The extensive simulation results show that AnDet effectively identifies anomalies in SDN-enabled cellular networks operating in B5G environments, achieving an impressive 97.2% accuracy in anomaly detection.**

***Index Terms*—Anomaly detection, Machine Learning, SDN, B5G networks, Explainable AI.**

I. Introduction

As cellular wireless networks progress towards the next generation of connectivity, popularly known as Beyond 5G (B5G), effective network management and enhanced Quality of Service (QoS) are becoming increasingly critical [1], [2]. Software Defined Networks (SDN) have emerged as a promising architecture for cellular networks, offering centralized control and dynamic resource allocation [3]. However, detecting and mitigating anomalies in these complex and scalable networks pose significant challenges [4]. Traditional rule-based approaches struggle to keep up with the dynamic nature and diverse range of anomalies. Consequently, integrating Artificial Intelligence (AI) and Machine Learning (ML) techniques has gained attention as a powerful solution for anomaly detection in SDN-Enabled B5G networks [5]. ML models can identify unexpected behavior by learning patterns, detecting deviations, and recognizing anomalous behaviors by analyzing the throughput and signal parameters of the User Equipment (UE). These parameters indicate delays in service, poor signal strength, and lower throughput, which often lead to service outages, resulting in poor QoS. Therefore, advanced algorithms are being developed to enhance QoS for delivering seamless services over future mobile networks.

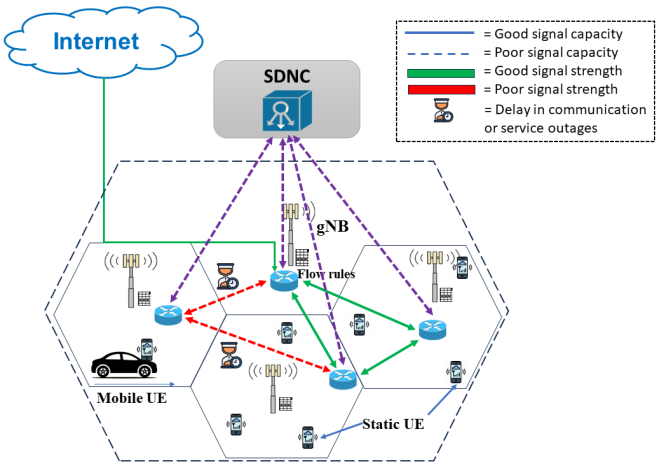

Fig. 1. Illustration of an SDN-Enabled B5G Multi-Cell Wireless Network

Motivation: Consider a scenario of SDN-enabled B5G multi-cell Wireless networks having many static and UEs. In this scenario, UEs are connected to the gNB. These UEs continuously communicate with the internet via an SDN-enabled B5G access network. Further, a change in the signal quality or an increase in capacity due to the mobility of users may result in communication delays. This situation leads to service outages, which result in discontinuity of services and poor QoS. As a result, the throughput is low in all the cases. Figure 1 depicts the scenario visually, demonstrating that as capacity increases, signal quality deteriorates and communication delays occur.

Therefore, to overcome these challenges and provide users with a seamless service experience. The convergence of B5G and SDN poses to be a potential solution. Furthermore, the scalability and adaptability of ML techniques make them well-suited for anomaly detection in B5G networks. The integration of Explainable AI (XAI) techniques helps to interpret the decisions and predictions which help in satisfying the regulatory requirements with an enhanced ML model, viz.

979-8-3503-7099-7/24/$31.00 ©2024 IEEE

TABLE I
COMPARISON TO THE STATE-OF-THE-ART

Solutions	Technique used	Adaptive AD	Unified	Focusing 5G	Near real-time	Designed For
F. L. Maimo *et al.* [6]	DL	✓	×	✓	✓	Traditional
H. A. Alamri *et al.* [3]	ML	×	×	×	×	SDN
L. Lei *et al.* [4]	Ensemble Learning	✓	×	✓	✓	SDN
A. R. Yeruva *et al.* [7]	ML	×	×	×	×	Traditional
S. Sinha *et al.* [5]	ML, Edge	✓	✓	×	✓	Traditional
V. S. Gawali *et al.* [1]	DL	×	×	✓	×	Traditional
AnDet	ML	✓	✓	✓	✓	SDN

debugging. Many studies relating to anomaly detection have been conducted but failed to address this use-case in an SDN environment. Insights to the literature have been furnished in the subsequent section II.

Contributions: In this research work, we make the following major contributions:

- **Development of Robust ML Model:** We have developed a robust ML model specifically designed for anomaly detection in SDN-enabled B5G multi-cell networks with an impressive accuracy of 97.2%.
- **Deployment of proposed model:** We have deployed the developed ML model in an SDN environment. Improving performance and reliability by identifying and mitigating anomalies, besides enhancing overall QoS.
- **Interpretation using XAI:** We have integrated XAI techniques in B5G cellular networks to interpret ML model decisions, fostering reliability and transparency.
- **Evaluation:** We have conducted large-scale simulation tests to validate our solution.

Organization: The rest of the paper is organized as follows: Section (II) is a review of the literature with a comparison to the state-of-the-art. Section (III) describes the proposed approach in conjunction with the system model. Section (IV) evaluates and discusses our final outcomes. Finally, in section (V), we conclude and outline the future scope.

II. Related work

Many approaches, including statistics, machine learning, and information theory, have been applied to the study of anomaly detection in the recent past. Few solutions for 5G (and beyond) rely on statistics [8], [9], such as outlier identification based on probability distributions, entropy-based algorithms, etc., while most of them rely on supervised or unsupervised ML schemes, such as clustering, classification, and so on [2]. Anomaly detection with ML can be divided into four categories: unsupervised, supervised, semi-supervised, and hybrid [10], [11].

Supervised models require that all data be labeled as normal or abnormal in order to detect anomalies [12]. Unsupervised anomaly detection, on the other hand, takes the opposite approach, requiring no training to analyze data and distinguish between normal and abnormal patterns [13]. In semi-supervised learning, however, different models predict an anomalous event using only normal data for training [11]. Service degradation and poor signal identification in mobile wireless networks are two areas where clustering algorithms have been used in previous research [13], [14]. To identify such anomalies in network data, Parwez *et al.* [13] use k-Means and hierarchical clustering. Performance loss owing to mobility issues was also postulated by Moysen *et al.* [14], who advocated for an unsupervised learning-based method.

A novel framework focused on proactively detecting such traffic anomalies is proposed by [2]. It includes a mechanism for automatically identifying various network traffic behaviors and forecasting network traffic behaviors for the time ahead, which could span several seconds. Recently, Maimó *et al.* proposed a new cyber defense architecture for 5G mobile networks that uses deep learning methods to extract features from network flows and examine anomalies in network traffic [6]. To mitigate the security challenges of network intrusion in an SDN-enabled 5G network, Lei *et al.* proposed an abnormal traffic detection algorithm based on ensemble learning [4] and experimented on abnormal traffic datasets. H. Alamri *et al.* proposed extreme gradient boosting (XGBoost) and other ML algorithms that use adaptive bandwidth mechanism and dynamic threshold technique to defend SDN controller targeted by DDoS attacks and improve the overall network performance [3]. Very recently, S. Sinha *et al.* provides a details taxonomy and discusses anomaly detection for 5G and edge computing [5]. The paper then explained how researchers implemented their anomaly detection model while considering edge nodes' computing, storage, and bandwidth limitations. In 2022, Yeruva *et al.* examined the differences and similarities of several ML-based network security anomaly detection systems and provided some insights for future research [7]. Finally, Gawali and Ranjan proposed an anomaly detection model based on deep learning for 5G networks [1]. This solution begins with feature extraction and then pre-processes the input data using a hybrid classifier (HC), including Deep Belief Network (DBN) and Bidirectional Long-Short-Term Memory (BiLSTM). However, these works are partly related to our research and target the security domain. A brief comparison to the state of the art is described in Table I.

Synthesis. The majority of the existing works [1], [3]–[7] lack real-time implementation of anomaly detection and mitigation in an SDN-enabled B5G environment. Rather, it focuses primarily on security aspects from a threat detection standpoint. In contrast, we propose a solution that directly

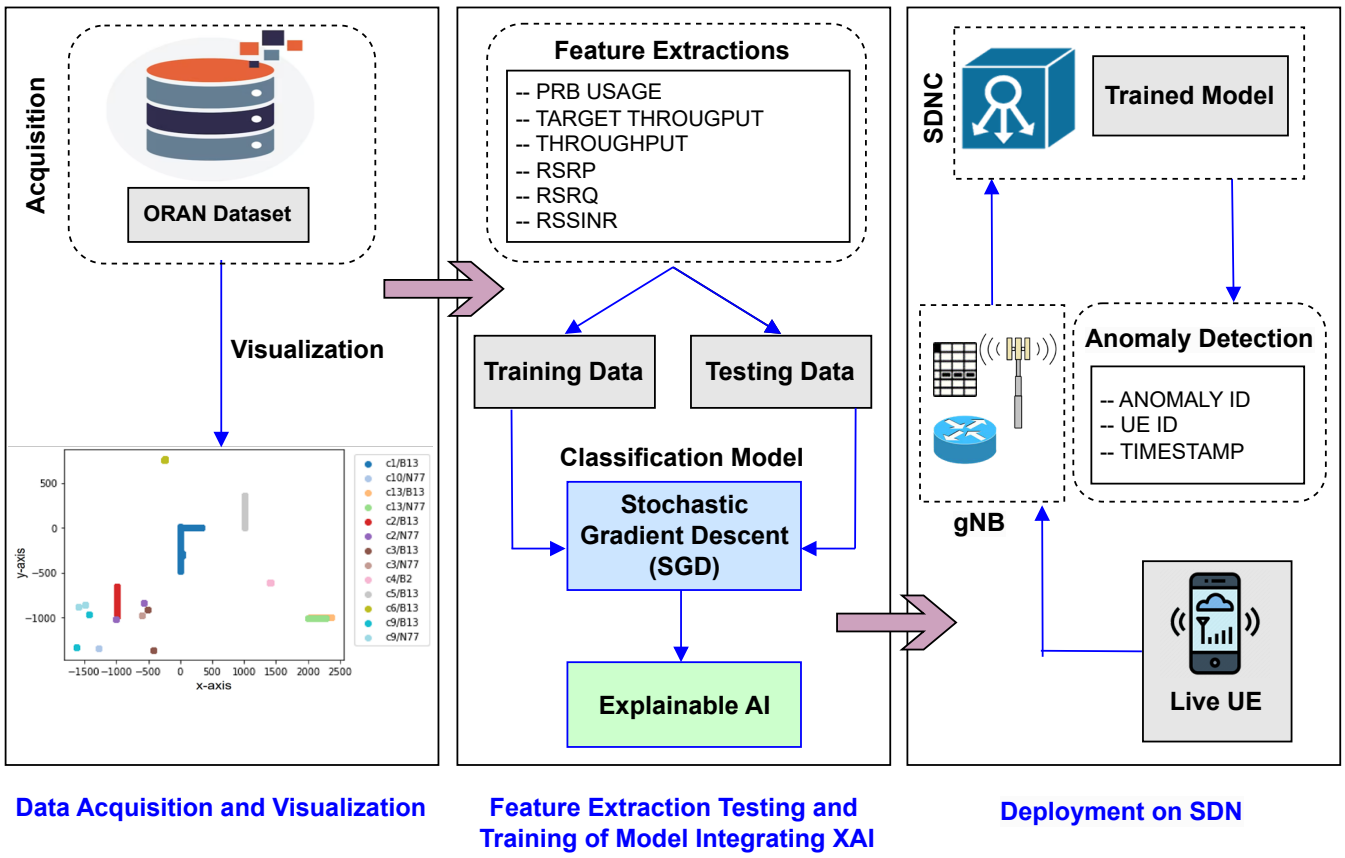

Fig. 2. A conceptual overview of the proposed solution

addresses challenges by emphasizing the enhancement of the SDN controller by incorporating ML-based models for improved QoS. A brief comparison of our proposed work to the state-of-the-art solutions has been shown in Table I.

III. Proposed Solution

This section presents an anomaly detection approach for SDN-Enabled B5G wireless networks. The study considers five neighboring cells connected through a central gNB. The system flow, as shown in Fig. 2, involves dataset gathering, feature extraction, model training and testing, and deployment in SDN. The approach aims to identify anomalies by extracting meaningful features from the collected data and training a model to differentiate normal and abnormal network behavior. The trained model can then be integrated into an SDN environment for centralized network management and control.

A. Data Acquisition

The dataset used in this research was obtained from the most recent release of the Anomaly Detection (AD) xApp, which is available in the official O-RAN repository [14]. It is made up of 10,000 simulated data samples gathered at 10 ms intervals between 11:42:33.695 and 11:43:06.815. The dataset contains a variety of features and attributes. Each data sample has its own du-id, which ranges from 1001 to 1006, 1009, 1010, and 1013. Every data sample has a timestamp, which indicates when the measurement was taken. The collection provides information about many cells labeled c1 to c6, c9, c10, and c13. The ue-id attribute defines the type of user equipment, which can be a car, a waiting passenger, or a train passenger.

The measures included in Signal Info are Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), and Reference Signal Signal-to-Interference plus Noise Ratio (RSSINR). Physical Resource Block (PRB) utilization and throughput information are detailed in Resources Info. Prelabeled anomaly instances with binary labels 0 and 1, indicating the absence and presence of anomalies, respectively. There are about 2568 samples of label 0 out of 10000 samples, indicating that about 25.7% data samples have anomalies.

To get a comprehensive understanding of the scenario, we have plotted the coordinates of the gNBs and UEs in the X-Y plane, leveraging clustering techniques. Upon careful examination, we have identified a prominent cluster of six cells, with one cell identified as the current cell and the other five as neighboring cells, inscribing the macro umbrella cell.

B. Anomaly Detection Model

The features used to develop the models were manually selected after checking various parameters, such as correlation and some visualizations. To compare all of the classification

models at once, we have utilized the *LazyPredictor* library [15]. Among various classifiers, the Stochastic Gradient Descent (SGD) classifier performs the best due to its rapid execution and excellent accuracy in anomaly identification. It operates by initializing model parameters like weights and biases and then performing iterative updates that involve calculating the gradient of the loss function for a subset (batch) of training data. Then, a learning rate scales the parameters in the opposite direction of the gradient. This cycle is repeated until the loss converges to zero or a predetermined stopping criterion is met. The dataset utilized for this was split between training and test sets in a 7:3 ratio so that we could evaluate the model on unknown data and prevent overfitting.

C. Explainable AI

Explainable AI (XAI) techniques, such as Local Interpretable Model-Agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP), increase transparency and human understanding of machine learning algorithms, particularly in anomaly identification and diagnostics [16]. This study leverages LIME XAI to comprehend the choice and applicability of the machine learning model proposed for anomaly detection in SDN-enabled B5G cellular networks. By integrating XAI, stakeholders and network operators gain valuable insights into detected anomalies, enabling proactive measures to optimize network performance. The incorporation of XAI techniques aligns with the motivation and contributions of the study, addressing the need for robust anomaly detection mechanisms in dynamic network environments. Transparent explanations of the model's decisions and influencing factors empower stakeholders to improve network security, reliability, and performance.

D. SDN-based Mitigation

SDN leverages flow rules to make communication happen. So, in context to the scenario, there are preinstalled flow rules as the UEs are continuously communicating to the internet. Further, a change in signal quality or mobility requires fresh flow rules. This is due to a handover event being triggered at the back-haul. So to maintain a continuous connection after anomaly detection, the flow rules need to be installed in the gNB of the best fit cell. To find out the best cell, we have considered the signal and data parameters from inside the same UE terminal are predicted rather than using metrics from each cell report. This helps reduce the constant data exchange in the control plane. With the scenario having five neighboring cells in addition to the serving cell, rather than executing a loop to extract five reports for each cell from the database, extracting the information about the cells from the UE report will mean saving up to 90% of the data that is being exchanged between each unique UE.

IV. Result and Discussion

In this section, we evaluated classification reports from various ML models and presented simulation results from the best model. The model's results have also been explained.

A. Performance of Anomaly Detection Models

The anomaly detection module depends upon various parameters such as accuracy, precision and recall, F1 score, and others. Accuracy is one of the most important and commonly used metrics for model performance evaluation. The overall classification accuracy of the SGD classifier is 97.2%. The comparison of the different models on the same dataset in Table II supports the implementation of SGD classifier in this dataset. The table demonstrates that the SGD classifier is an excellent choice and performs better in comparison to other models. Table III presents the performance report, including evaluation metrics such as accuracy, precision, recall, and F1-score. These metrics provide an assessment of the classification task's effectiveness.

TABLE II
PERFORMANCE OF DIFFERENT ML MODELS WITH THE SAME DATASET

Model	Accuracy	Training time	Prediction time
SGD	97.2 %	15.6 ms	0.9 ns
Random Forest	97.3 %	1.31s	1.4 ns
LGBM	97.1 %	547 ms	141ms
XGBoost	97.1 %	1.03s	1.5 ns

TABLE III
SGD PERFORMANCE REPORT

Fault Type	Precision	Recall	F1-score	Accuracy
Anomaly	96.36%	1.00%	98.15%	97.2%
No Anomaly	1.00%	89.10%	94.23%	

B. Deployment on SDN

Conceptualizing the scenario in an SDN environment, based on the data we have collected, we have visualized an experimental network topology. This topology is illustrated in Fig. 3. In this design, we have implemented an umbrella cell

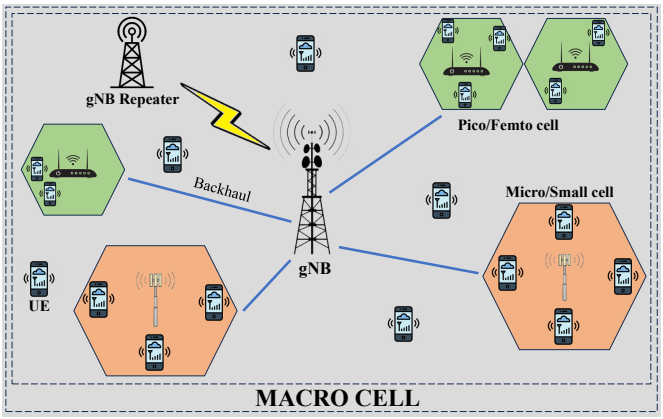

Fig. 3. Proposed experimental topology

approach, which aims to optimize the cell coverage in the network. The primary objective of the umbrella cell approach is to maximize the coverage area for mobile users while ensuring efficient connectivity for slow or static users within

a narrower region. By strategically deploying cells and access points, we can create a network architecture that caters to the diverse needs of different user types. Also, we can observe a high-level representation of the experimental network topology. The network consists of multiple cells, each represented by a dotted region. These cells are strategically placed to ensure comprehensive coverage throughout the target region. The larger cells in the figure represent the umbrella cell. These cells are designed to cover extensive areas, providing seamless connectivity for mobile users constantly moving within the network. The umbrella cells are equipped with high-capacity access points, enabling them to handle a larger number of concurrent connections and higher data traffic.

It is worth noting that the specific implementation and configuration of the network topology may vary based on the actual deployment requirements, available resources, and network demands. However, the conceptualization presented here serves as a foundation for designing an SDN environment with an umbrella cell approach to maximize cell coverage and optimize network connectivity for diverse user requirements.

We have simulated the scenario in consideration in the *Mininet-WiFi* [17] simulation tool in VM, with practical conditions, where a total of 10 UEs are moving in a random direction across a range of 1 km radius. This consists of 6 gNBs operating at unlicensed frequencies, i.e., 2.4 GHz and 5.8 GHz, which differ in radio access technologies, Tx power, ranges, etc. Further, UEs were simulated as STAs and the gNBs as APs, configuring the signal ranges as represented in Fig 4. Also, wireless media was replicated using *wmediumd*

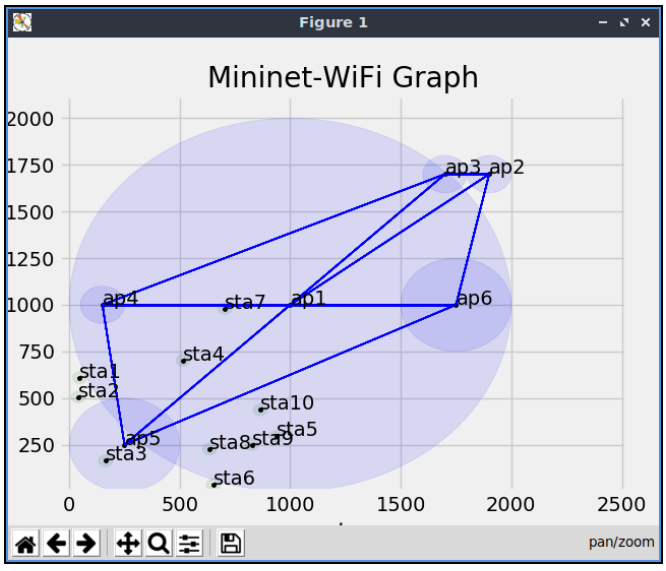

Fig. 4. Simulation of scenario in Mininet-WiFi

[18], allowing it to resemble the physical characteristics of the channel. The interference model enhanced the simulation by calculating real signal parameters, which helped to accurately predict the anomalies while providing a global view of the network. Further, a central SDN controller was extended as a function of the RYU controller. A summary of simulation parameters is indicated in Table IV.

TABLE IV
SUMMARY OF SIMULATION PARAMETERS CONFIGURED FOR THE PERFORMANCE EVALUATION OF THE SCENARIO

Parameter	Value
No. of macro gNB / Tx power / Range	1 / 91 dBm / 1000 m
No. of pico gNB / Tx power / Range	2 / 60 dBm / 250 m
No. of femto gNB / Tx power / Range	3 / 41 dBm / 100 m
Tx power of UEs	14 dBm
Controller protocol	OpenFlow
Propagation model	Log distance
Mobility model	Random
Number of SDN controllers	1
Number of hosts	1
Number of UEs	10
UE Connectivity	TCP
Class of network	A (10.0.0.0/8)
Containerization	VM

Some stations were given mobility to mimic the dynamic users. A random speed was assigned to resemble a practical scenario. Subsequently, we have captured logs of different events and also the performance metrics for evaluation of the deployed solution.

C. XAI-based SGD Model Interpretation

For a better understanding of the predictions given by the black-box ML model, it is important to get a clear understanding of the effects of each input characteristic on each target class. This was accomplished using Local Interpretable Model-Agnostic Explanations (LIME), an XAI methodology implementation. Local instances of the test data can be declared for each classification of labels and supplied in a LIME explanation to demonstrate the variances and particular significance of the characteristics.

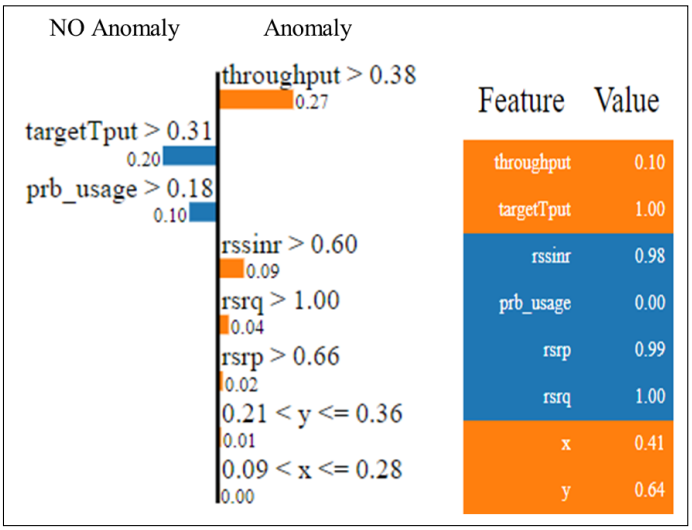

Fig. 5. LIME explanation plot for anomaly and no anomaly situation

The LIME explanation for the anomaly and the absence of an anomaly condition are shown in Fig. 5. It implies that

throughput plays a significant role in determining the anomaly state, whereas *targetput* (i.e., target throughput) and *prbusage* do the same for the safe or absence of anomalies. This would imply that these three characteristics actively influence whether a data sample is safe. This may also be seen in Fig. 6, which displays each feature's relevance and clarifies that these three features are of great priority.

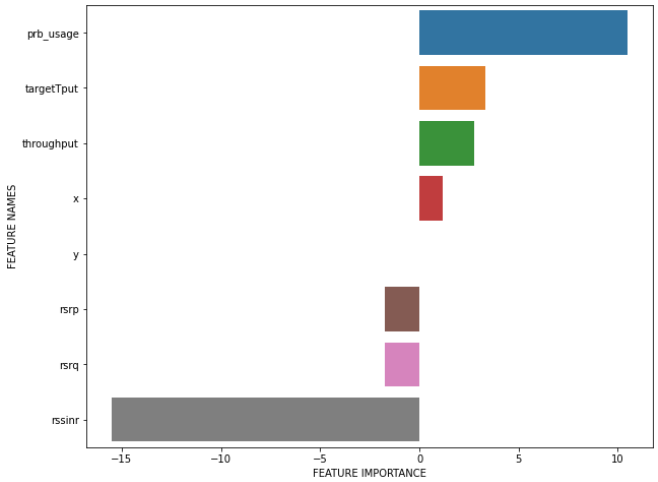

Fig. 6. Feature importance plot showing relative relevance of different features

V. Conclusion and Future Scope

This research work introduces a framework to enhance the capabilities of SDN controllers in B5G multi-cell networks, specifically focusing on anomaly detection and service improvement. By integrating advanced algorithms, the framework enables effective monitoring of network traffic, facilitating the identification of anomalies and triggering appropriate actions for mitigation. Experimental evaluations validate the framework's efficacy in accurately detecting anomalies and optimizing network services, demonstrating its potential for ensuring the smooth operation of complex B5G networks. The proposed solution leverages SDN technology to address challenges associated with network behavior, such as security breaches, performance degradation, and congestion. Future work will involve testing the proposed machine learning models with larger and live datasets to assess the framework's performance in more dynamic network environments. Moreover, further experiments will be conducted to evaluate the solution from a network management perspective, considering its impact on network operations, maintenance, and overall performance.

Acknowledgment

This research work is funded by IIITB COMET Foundation, established under the National Mission on Interdisciplinary Cyber-Physical System (NM-ICPS), Department of Science and Technology (DST), Government of India (GoI).

This work is also supported in part by DoD Center of Excellence in AI and Machine Learning (CoE-AIML) under Contract Number W911NF-20-2-0277 with the U.S. Army Research Laboratory, National Science Foundation under Grant No. 2219742 and Grant No. 2131001.

References

[1] V. S. Gawali and N. M. Ranjan, "Anomaly detection system in 5g networks via deep learning model," *International Journal of Wireless and Mobile Computing*, vol. 24, no. 3-4, pp. 287–302, 2023.

[2] N. Koursioumpas, L. Magoula, S. Barmpounakis, and I. Stavrakakis, "Network traffic anomaly prediction for beyond 5g networks," in *2022 IEEE 33rd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE, 2022, pp. 589–594.

[3] H. A. Alamri, V. Thayananthan, and J. Yazdani, "Machine learning for securing sdn based 5g network," *Int. J. Comput. Appl*, vol. 174, no. 14, pp. 9–16, 2021.

[4] L. Lei, L. Kou, X. Zhan, J. Zhang, and Y. Ren, "An anomaly detection algorithm based on ensemble learning for 5g environment," *Sensors*, vol. 22, no. 19, p. 7436, 2022.

[5] S. Sinha, D. S. Tomar, and R. Pateriya, "Anomaly detection for edge computing: A systematic literature review," in *AIP Conference Proceedings*, vol. 2705, no. 1. AIP Publishing, 2023.

[6] L. F. Maimó, Á. L. P. Gómez, F. J. G. Clemente, M. G. Pérez, and G. M. Pérez, "A self-adaptive deep learning-based system for anomaly detection in 5g networks," *Ieee Access*, vol. 6, pp. 7700–7712, 2018.

[7] A. R. Yeruva, P. Chaturvedi, A. Rao, S. C. DimriL, C. Shekar, and B. Yirga, "Anomaly detection system using ml classification algorithm for network security," in *2022 5th International Conference on Contemporary Computing and Informatics (IC3I)*. IEEE, 2022, pp. 1416–1422.

[8] K. Yang, H. Ma, and S. Dou, "Fog intelligence for network anomaly detection," *IEEE Network*, vol. 34, no. 2, pp. 78–82, 2020.

[9] A. Pavlidis, G. Sotiropoulos, K. Giotis, D. Kalogeras, and V. Maglaris, "Nfv-compliant traffic monitoring and anomaly detection based on dispersed vantage points in shared network infrastructures," in *2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft)*. IEEE, 2018, pp. 197–201.

[10] M. E. M. Cayamcela and W. Lim, "Artificial intelligence in 5g technology: A survey," in *2018 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2018, pp. 860–865.

[11] T. Sundqvist, M. Bhuyan, and E. Elmroth, "Uncovering latency anomalies in 5g ran-a combination learner approach," in *2022 14th International Conference on COMmunication Systems & NETworkS (COMSNETS)*. IEEE, 2022, pp. 621–629.

[12] J. Lam and R. Abbas, "Machine learning based anomaly detection for 5g networks," *arXiv preprint arXiv:2003.03474*, 2020.

[13] M. S. Parwez, D. B. Rawat, and M. Garuba, "Big data analytics for user-activity analysis and user-anomaly detection in mobile wireless network," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 2058–2065, 2017.

[14] J. Moysen, F. Ahmed, M. García-Lozano, and J. Niëmela, "Unsupervised learning for detection of mobility related anomalies in commercial lte networks," in *2020 European Conference on Networks and Communications (EuCNC)*. IEEE, 2020, pp. 111–115.

[15] Shankar Rao Pandala, "Lazy Predict," Python Software Foundation, 2019. [Online]. Available: https://pypi.org/project/lazypredict/

[16] V. Margot, "A brief overview of methods to explain ai (xai)," *Towards Data Science*, 2021.

[17] Ramon dos Reis Fontes, "Mininet-WiFi: Emulation Platform for Software-Defined Wireless Networks," Github, 2023. [Online]. Available: https://mininet-wifi.github.io/

[18] Bob Copeland, "Wmediumd: Wireless medium simulation tool," Github, 2023. [Online]. Available: https://github.com/bcopeland/wmediumd