Benchmarking of Anomaly Detection Techniques in O-RAN for Handover Optimization

Zineb Mahrez°, Maryam Ben Driss°, Essaid Sabir°•, Walid Saad*, and Elmahdi Driouch•
°NEST Research Group, LRI Lab, ENSEM, Hassan II University of Casablanca, Morocco
•Department of Computer Science, University of Quebec at Montreal, Montreal, H2L 2C4, Quebec, Canada
*Department of Electrical and Computer Engineering, Wireless@VT, Virginia Tech, VA, USA
Emails: zineb.mahrez@ensem.ac.ma; maryam.bendriss@ensem.ac.ma; essaid.sabir@ieee.org; walids@vt.edu; driouch.elmahdi@uqam.ca

***Abstract*—With today's proliferation of IoT and real-time applications, it has become crucial to properly handle traffic, optimize the quality-of-service (QoS) of wireless networks and design novel approaches to enable reliable communications with bounded latency and high throughput for future wireless services. In order to meet these stringent QoS requirements, there has been a recent surge in research that investigates a deep restructuring of the Radio Access Network (RAN). In particular, the open radio access network (O-RAN) framework promises to deliver flexible, scalable, and agile solutions for improving the handover process of a moving user equipment (UE), by considering factors related to the requirements of applications in terms of QoS, traffic load, signal quality and the diversity of multiple access technologies or radio frequencies of the environment. Building on this basis, this paper investigates the handover process of a moving vehicle, by exploring and comparing the prediction accuracy results of different machine learning (ML) techniques used for anomaly detection. In particular, the structure of one of the O-RAN modules is presented. This module relates to a traffic steering application, specifically designed and used to detect anomalies within the network. Several ML techniques are then implemented in the O-RAN traffic steering module to predict the handover. The results pertaining to the comparison of the implemented ML techniques show that the random forest algorithm gives the highest accuracy (up to 98%), which helps boosting the handover process.**

***Index Terms*—5G, Anomaly Detection, Handover, Artificial Intelligence, Machine Learning, Open RAN, RIC, xApp.**

I. Introduction

The evolution of mobile communication technologies and the increasing demand of real-time applications have motivated the deployment of flexible protocols, such as those based on artificial intelligence (AI), to manage and operate wireless networks [1], [2]. However, in order to deploy such new protocols, it is necessary to rethink the way in which the radio access network (RAN) is designed. In particular, existing RAN systems are limited in terms of supporting the varied requirements of 5G applications. These limitations include a lack of data-driven functionality, a lack of support for the deployment of artificial intelligence, limited configurability, closed proprietary systems, as well as the interoperability issue between RAN components supplied by different vendors. For instance, traditional RAN systems are not able to handle the high data rates required by modern applications such as video streaming and online gaming that has very stringent requirements in regard to throughput and bandwidth demand. Additionally, RAN structure is not able to provide the low latency needed for real-time applications such as voice and video calls. Furthermore, it's unable to provide the flexibility needed to accommodate a wide range of services, user needs and changing traffic patterns. These limitations are partly due to the lack of standardization of interfaces that allow data collection from the base stations [3]. Clearly, there is a need for new RAN standards that are more flexible and agile, and that could accommodate AI models and algorithms. For instance, the lack of openness in the RAN has been identified as a major bottleneck in maximally utilizing network virtualization. In response to all these needs, many operators such as AT&T, China Mobile and many others have joined forces to found an alliance in 2018 to promote the concept of an intelligent and open RAN that is not tied to proprietary functions. This alliance is called the open radio access network, or O-RAN [4]. Its mission is to re-shape the industry towards more intelligent, virtualized network elements, white-box hardware, standardized and open interfaces as shown in Fig. 1. In this sense, O-RAN emerged as a new modern wireless network model, that can respond to the requirements of 5G and beyond, by (i) opening interfaces between the base-band units (BBUs) and the remote radio units (RRUs), which offers flexibility, scalability, and agility of RAN networks thus enabling multi-vendor deployments to achieve an optimal solution, (ii) enabling network automation, (iii) deploying intelligence, and (iiii) enhancing RAN performance through virtualization. One of the most important part of the O-RAN framework is the radio intelligent controller (RIC), that plays the role of a traditional network controller in charge of orchestrating network functions, by performing a proactive maintenance of traffic network. Also called as open smart controller, it consists of embedding intelligence into the radio controller, along with splitting user and control planes, which will succeed in improving the traditional radio resource management functions (RRM) [5] and [6]. It is composed of two layers, the near real-time RIC and the non real-time RIC. On the one hand, the near real-time RIC operates in near-real-time, in the time-frame of 10 ms to 1s and is responsible for RAN control and optimization, it hosts third parties xApps such as machine learning

This research was supported by the U.S. National Science Foundation under Grant CNS-2225511.

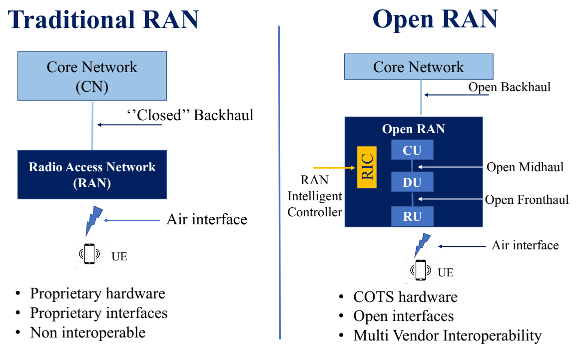

Fig. 1. Traditional RAN vs. Open RAN

(ML) xApp, self optimization network xApp. On the other hand, the non real-time RIC is part of service management and orchestration (SMO). It allows operators to flexibly configure the desired optimization policies, utilize the right performance criteria, and leverage machine learning to enable intelligent and proactive traffic control. It operates in a timeframe more than 1s. It uses RAN data analytic to perform policy guidance and training of machine learning models. It is called non real-time because it operates on offline mode in a duration greater than one second [7]. The state-of-the-art analysis reveals that very few research works have thus far developed concrete solutions for deploying machine learning in O-RAN networks [3], [6], [8]–[10]. In addition, the integration of the RIC to improve smart mobility applications particularly vehicular-to-everything (V2X) has not been considered yet, and still not available for deployment as an xApp within O-RAN software. The main contribution of this paper is the investigation of handover management in O-RAN. We used the traffic steering module that incorporates anomaly detection and we evaluated and compared the performance of various ML algorithms for handover prediction, including local outlier factor, isolation forest, robust covariance, logistic regression, naive bayes, SVM and random forest. Given the structure of our database, our results demonstrate that the random forest used for anomaly detection in the traffic steering module of O-RAN can significantly improve handover performance and outperform other methods.

The rest of this paper is organized as follows. Section II presents a background and related works to introduce anomaly detection techniques on handover management. This section also describes the different third party applications used in the O-RAN environment for anomaly detection. Section III presents a detailed analysis of the network data using anomaly detection. Section IV presents the corresponding results and finally, section V provides concluding remarks and future works.

II. Background and Related Works

To bridge the gap between the machine learning algorithms and their application in O-RAN, we first introduce previous works focusing on machine learning algorithms for detecting anomalies, and highlight related works in this area, then we provide a detailed explanation of how anomaly detection can be applied in O-RAN environment.

A. Machine learning algorithms for anomaly detection

Anomaly detection (AD) is an unsupervised machine learning technique in charge of detecting abnormal data (outliers) that deviate from the norm (inliers). These types of data are called anomalies and are classified into 3 different categories [11]:

- **Outliers:** Refer to short or small anomalous patterns that appear in a non-systematic way in data collection (for instance communication errors).
- **Change in events:** It happens when a sudden change has occurred (for instance extreme weather conditions).
- **Drifts:** Are characterized by the slow, unidirectional, and long-term change in data, for instance, a fault in a sensor.

In the context of O-RAN, AD algorithms identify abnormal behavior of network data such signal power or communication between nodes. Several approaches could be used to perform anomaly detection, such as machine learning algorithms based on (1) decision tree, classification or statistical models, (2) clustering such as K-means where anomalies do not belong to a cluster, (3) deep reinforcement learning, specifically used to handle high dimensional data and multiples features.

1) Classification:

Classification for anomaly detection assigns a predefined label to a given input feature, it involves predicting and detecting irregularities in real-time and helps triggering proper actions such as handover. While various classification ML algorithms are available, we have chosen to focus on logistic regression, naive bayes, support vector machines (SVM), and random forest because of their performance in detecting irregularities and their applications to the specific context of network data.

Logistic Regression

Logistic regression is a supervised ML classification algorithm that predicts the probability of occurrence of an event by learning the relationship between the features and mapping inputs data to a probability outputs with values between 0 and 1. Let's consider X to be the input data and θ be the parameter we want to train and optimize, the probability is then expressed by the Sigmoid function as following :

$$h_\theta(X) = \frac{1}{1 + e^{-\theta X}} \tag{1}$$

Naive Bayes

Naives Bays is a supervised learning and a probabilistic classifier that belongs to a family classifier algorithms based on Bayes' Theorem (See equation 2), and where every pair of features being classified is assumed to be independent of each other. Naive Bayes classifier calculates the probabilities and selects the highest probability outcome.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2}$$

Support Vector Machine (SVM)

SVM is a supervised machine learning used for classification, regression and anomaly detection. The role of this algorithm is to define a decision boundary (also called "a hyperplane"), consisting of a subset of training points: support vectors. The hyperplane helps classifying data points by splitting data into two main categories, its dimension depends on the number of features. Many hyperplanes could be selected, however an effective classification of SVM algorithm requires a maximum margin distance between data points of both classes. During the training, the SVM learns to make a prediction for a new data point, by measuring the distance to each of the support vectors.

Random Forest (RF)

Random forest is considered to be one of the most important supervised learning algorithm and is based on four steps. In the first step, we consider a dataset of k records, where n random records are retrieved. The second step consists of building individual decision trees for each sample. In step 3 each decision tree will generate an output. And in step 4, final output is considered based on majority voting or averaging for classification and regression respectively.

Given the significant volume of data collected from the RAN, the classification labeling process of RAN anomalies might be challenging and costly [12]. In this sens, other alternatives can be considered, such as clustering or DRL as discussed next.

2) Clustering:

This subsection describes the clustering algorithms used for anomaly detection and discuss its application in O-RAN environment.

Local Outlier Factor (LOF)

LOF is an unsupervised machine learning technique based on density and nearest neighbors algorithm, useful for anomaly detection. The application of LOF is based on the four following concepts:

- K-distance and K-neighbors: K-distance is the distance between a given point X_i to its K^{th} nearest neighbor, defined and computed according to the value of the hyper-parameter k that represents the number of neighbors. K-neighbors denoted by $N_k(X_i)$ represent a set of points that belong to the circle of radius K-distance.
- Reachability distance (RD): The RD refers to the distance of travel from a particular point X_i to its neighbor point X_j. RD is defined as the maximum of the distance between these two points and the K-distance. For instance, in the case of the Euclidean distance, we consider X^k to be the k^{th} nearest neighbor of X_i, then the RD from X_i to X_j is expressed as following:

$$RD_k(X_i, X_j) = \max(|X_i - X^k|, |X_i - X_j)|) \tag{3}$$

- Local reachability density (LRD): We consider $X_i \in \mathcal{X}_i^l$ with $l \in [1, n]$. The LRD is equal to the inverse of the average RDs from $X_i^{(l)}$ to X_i [13], as expressed by equation 4. In a practical way, the LRD states that the more the average RD is (X^k from X_i), less density of points are present around a X_i.

$$LRD_k(X_i) = \left(\sum_{i=1}^{k} \frac{RD_k(X_i^{(l)}, X_i)}{N_k(X_i)} \right)^{-1} \tag{4}$$

- Local Outlier Factor (LOF):
  The LOF represents the ratio of the LRD of a given point X_i to the LRD of its K-neighbors. Based on the LOF score, if $LOF > 1$, that means X_i is an outlier.

$$LOF_k(X_i) = \frac{1}{|N_k(X_i)|} \cdot \sum_{X_j \in N_k(X_i)} \frac{LRD_k(X_j)}{LRD_k(X_i)} \tag{5}$$

Robust Covariance

The Robust Covariance is an unsupervised outlier detection based on probabilistic distribution. Let's consider a point x_i following a probabilistic distribution of mean μ and covariance matrix $cov[x_i]$, then the deviation x_i from μ is known as Mahalanobis distance d_i, expressed in the following:

$$d_i(\mu, cov[x_i]) = \sqrt{(x_i - \mu)^T cov[x_i]^{-1} (X_i - \mu)} \tag{6}$$

In this section, we discussed the use of classification and clustering algorithms for anomaly detection (AD) in general. In the context of O-RAN, it offers an effective way to identify anomalies in terms of volume of data, disaggregated architecture, real-time detection, and high sacalability. In typical RAN, the AD algorithms can be applied only to one data source whereas in O-RAN, AD algorithms are applied to a significant volume of network data collected from a heterogeneous environment, and from multiple data sources. For instance, authors in [3] proposed a clustering solution to reduce control plane latency by decreasing the number of inter-controller interactions and described how this proposed framework could be integrated in the O-RAN architecture. They applied the proposed framework to study the reduction of prediction error for 600,000 user traffic in 650 base stations. In terms of architecture, AD algorithms are processed in a single location centralized in the typical RAN. O-RAN offers the flexibility of AD algorithms to be implemented in different components, which allows the detection of irregularities in real-time and communicate with the RIC to execute the handover. Furthermore, O-RAN is designed to be able to support a large number of connected devices and users, unlike typical RAN that can not satisfy high scalability requirements.

3) Deep reinforcement learning:

Intelligent handover (HO) management in wireless systems has been a topic of significant research interest in recent years [14]–[16]. For instance, in [14], the authors deployed recurrent neural network (RNN) algorithm to predict the movement of traffic of subscribers towards a particular cell in the heterogeneous 5G mobile networks. On another context in [15], the authors proposed a dual stacked RNN with Long-short term memory (LSTM), that help vehicles predict handover points based on their interactions with fog computing nodes. With the advent of autonomous vehicles, handover management becomes more crucial as it is closely related to user safety, and thus requires a real-time application. Many researchers attempted to solve the handover problem for autonomous vehicles, using various solutions, including the LTE hybrid access proposed in [16], that consists of using a combination of LTE multiple operators. As part of the improvement of the innovation, and in order to enhance the performance of the handover management process, operators have been advocating for multi-vendor interoperability of RAN elements, which is now possible with O-RAN. To successfully operate IoT services such as connected autonomous vehicles, a wireless system must simultaneously deliver high reliability, low latency, and high data rates, for heterogeneous devices, across uplink and downlink. A number of recent works have investigated the various capabilities of O-RAN in handling handovers [**?**], [6], [8], [9]. In [6], authors propose a workflow of ML implementation in O-RAN through the RIC by using Acumos (a linux foundation open-source project that is interoperable with O-RAN architecture). Autonomous and self-optimizing networks using O-RAN are considered in [8], where authors proposed a data driven closed loop control to gain in spectral efficiency and buffer occupancy. Additionally, in [9], authors proposed an algorithm that uses automatic neighbor relation (ANR) xApp in O-RAN, that is based on random forest decision tree classifier, in order to optimize the neighbor cell relation table through the minimization of handover failures. Another example of predicting occurrences of handovers was studied in [**?**], in which the authors identified potential congested cells using a recurrent neural network (RNN) model and long short-term memory (LSTM). Overall, these study cases have contributed to a better understanding of the capabilities and potential benefits of O-RAN, however to the best of our knowledge, we found no prior research that benchmarked anomaly detection ML techniques in the context of O-RAN. Therefore, we present an overview process that aims to fill this gap in the literature.

In order to effectively apply ML algorithms of anomaly detection in the O-RAN environment, vendors and operators developed a specific framework called 'traffic steering' module available in O-RAN software, which is detailed in section II-B. This module is based on isolation forest which is an unsupervised machine learning algorithm used for anomaly detection. It 'isolates' data points by building an isolation tree, that randomly select a feature from the dataset and then selecting a split value, in charge of dividing the data points into two sub-trees. In the following section, we explore the internal architecture of traffic steering that may use various machine learning algorithms for anomaly detection.

B. Application of anomaly detection in O-RAN through the traffic steering module

"Traffic steering" is a process of selecting and directing a user equipment (UE) to the most convenient base station with the best possible QoS in a wireless communication network. In classical traffic steering processes, the selection operation was based on the movement of the device and the average values of signal power. In addition to this, all connected devices and applications were treated equally. This represents a challenge in heterogeneous networks (5G) where different types of cells are deployed. Furthermore, new emerging applications have different requirements while being deployed in the same device. This is the case of V2X critical application such as forward collision warning system that needs to change a cell earlier due to specific requirement of QoS, whereas, a non critical application such as infotainment doesn't have stringent QoS requirement. This leads us to the main question "how to design different strategies for different types of users". In this case, O-RAN enables us to use machine learning techniques to enable this flexibility and build intelligent traffic steering control by using new ready-to-use applications called xApps. One of the proposed xApps is the traffic steering xApp, used to enhance network performance by balancing the load across cells and optimize the handover process with the help of intelligent algorithms such as anomaly detection algorithm. The traffic steering xApp comprises of 5 open source sample xApps. Each xApp has a specific role as described in fig. 2 and in the following [17], [18]:

(1) KPIMON xApp: is a KPI monitoring xApp that collects data and metrics received by the user such as:

- RSRP: which is the reference signal received power and represents the average received power without noise and interference.
- RSSI-NR: that refers to the adjacent channel interference including serving and non-serving cells.
- RSRQ: which is the reference signal received quality metric represented as the ratio between RSSI and RSRP.

These metrics are collected from the O-RAN distributed unit (O-DU) which is responsible for real-time processing of radio scheduling information and beamforming, and the O-RAN centralize unit (O-CU) that controls the function of numerous DUs over the midhaul interface, through E2 nodes (E2 nodes represent the interface between the RAN and the base station). The data are then stored in a time series database (Influx database).

(2) Anomaly detection (AD) xApp: The AD xApp is in charge of detecting anomalous signal metrics that serve UEs within the network. The xApp retrieves the anomalous UE data from the Influx database and performs ML training. Currently, in the O-RAN software, the anomaly detection is based on "isolation forest model" to detect anomalous UEs. The type of anomaly or degradation as well as the data about the detected UEs are sent to the traffic steering xApp.

(3) Traffic steering (TS) xApp: Based on the list received from the AD xApp, the TS xApp starts the prediction operations on UE throughput. The results are then transmitted to the non real-time RIC, which is in charge of defining and providing policies to the near real-time RIC over A1 policy (A1 interface enables policy-driven guidance from the non-RT RIC to the near-RT RIC), in order to guide the behavior of traffic steering xApp. As illustrated in Fig. 2, there are four possible policies stated by A1 policy: SHALL, PREFER, AVOID or FORBID. For instance, in Fig. 2, A1 policy states that vehicle B should "AVOID" cell M2 and "PREFERS" to

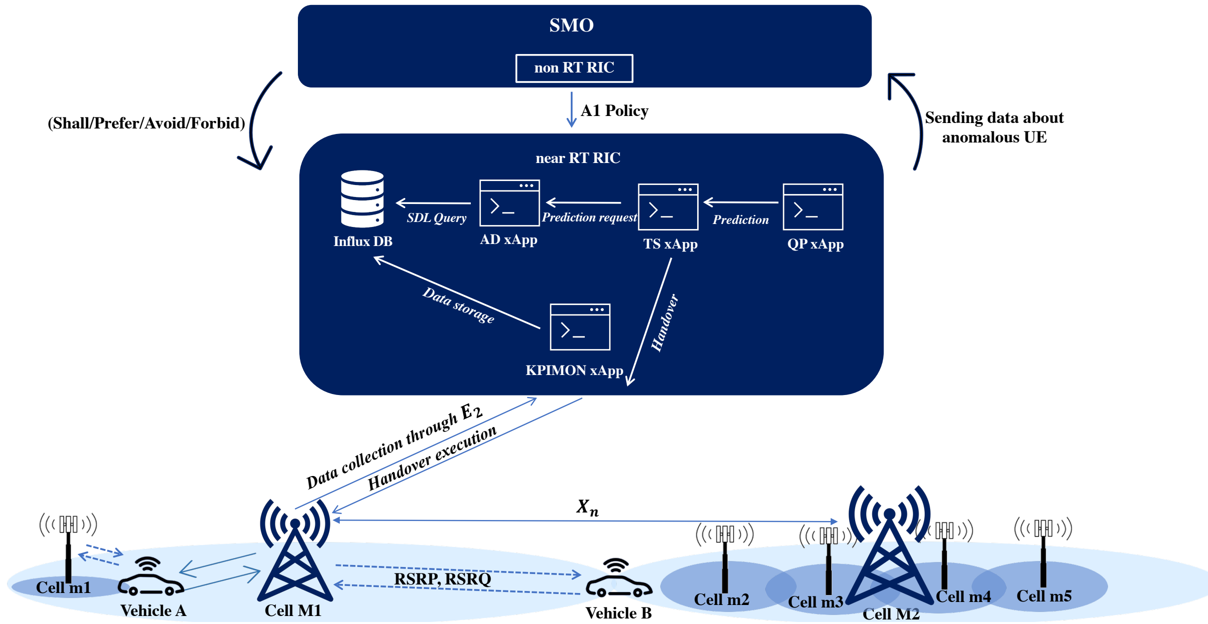

Fig. 2. Traffic steering process for handover operations.

connect to cell m2 which is a microcell able to deliver a high QoE signal compared to macrocell M2.

(4) QoE predictor (QP) xApp: QP xApp uses machine learning algorithm to predict throughput of the UE and sends the outputs to the TS xApp. The near real-time RIC executes these policies, to switch the connection to another cell that has a maximum throughput, and then sends control messages to E2 nodes to execute the handover.

These xApps are crucial elements in the O-RAN architecture and can significantly enhance the performance of handover by providing real-time monitoring and visualization of network KPIs, raising alerts in the presence of anomalies, improving network resource utilization and predicting end-user QoE.

III. Data analysis and simulation setup

Our work focused on using O-RAN to optimize the handover process of a user equipment moving from a cell to another cell and records signal strength measurements (RSRP, RSRQ) from the surrounding cells every second. The dataset contains 10,000 observations and 5 features. O1 interface records a set of standardized events in NR gNBs and collects data about UE movement pattern and distributes data across the entire O-RAN infrastructure. The dataset can be accessed from the official O-RAN github [4]. There main characteristics are :

- 25% of abnormal users.
- 9 cells and 5 neighbor cells.
- UE-ID refers to car, waiting passenger, train passenger.
- Signal strength parameters: [RSRP, RSRQ, RSSINR], 'PRB usage', 'Throughput'.

The input features used in the data analysis are related to measurement reports and UE performance statistics:

- RSRP metric has a range of: [-44dBm (excellent), 140dBm (No signal)]. In our dataset, the maximum and minimum values of RSRP are:[-150.37dBm, -56.81dBM].
- RSSI-NR: In our dataset, the maximum and minimum values of RSSI are:[-13.15 dBm, 65.41 dBm].
- RSRQ: In our dataset, the maximum and minimum values of RSRQ are:[-70.79 dBm, -10.79 dBm].
- Throughput refers to the actual measure of data successfully transferred from source to destination.
- Physical resource block (PRB) which is the smallest unit of scheduling block that a gNB assigns for transmission.

The first step is to apply data preprocessing to reduce the negative influence of irrelevant data or values that might be missing in the dataset. In our case, the preprocessing phase is based on the same approach used by O-RAN software. It is mainly based on the following steps:

- Dropping irrelevant data for predictions which include the UE ID, the UE category and timestamp.
- Checking and dropping high correlation parameters.
- Normalizing data to bring all parameters in same scale.
- Filtering the numeric data types.
- Dropping data that have NAN values.

Fig. 3 shows the mean and standard deviation of each feature. There is a high PRB utilization with an average greater than 93% of the cell serving anomalous UEs. Poor signal measurements (RSRQ) and poor network throughput are also noticed for abnormal users.

IV. Simulation results and analysis

We applied different ML techniques to predict anomalous users and then minimize the handover failure, as well as comparing the performance and scores of each ML technique

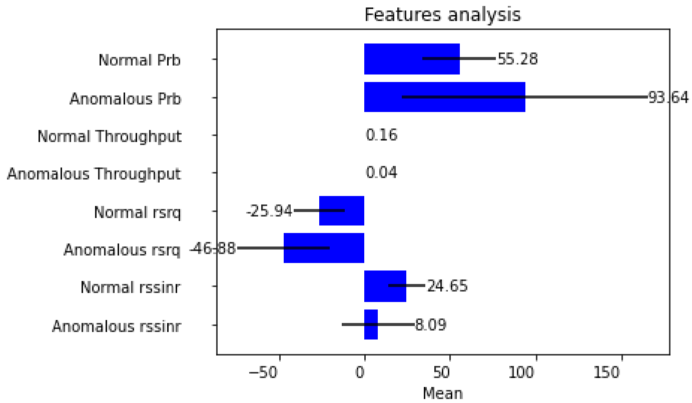

Fig. 3. Features analysis of normal users V.S. abnormal users.

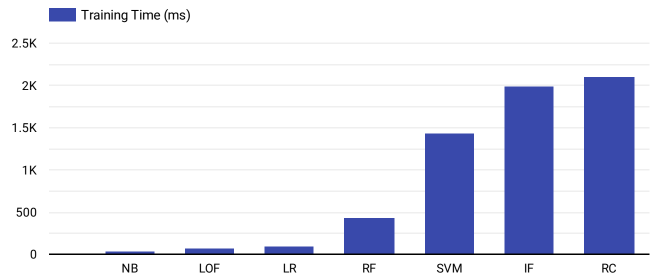

Fig. 4. Comparison of ML techniques in terms of training time (in ms).

TABLE I
DIFFERENT ML TECHNIQUES APPLIED TO MOBILITY DATA USING O-RAN

ML technique	Accuracy	Precision	Recall	F-Score	RMSE
Local Outlier Factor	0.7404	0.3627	0.0144	0.0277	0.50
Isolation Forest	0.9118	0.8010	0.8734	0.8356	0.29
Robust Covariance	0.8322	0.9504	0.3656	0.5281	0.40
Logistic Regression	0.1349	0.1386	0.4544	0.2124	0.93
NAIVE BAYES	0.8651	0.8850	0.5455	0.6750	0.36
SVM	0.8133	0.6657	0.5482	0.6013	0.43
Random Forest	0.98	0.9760	0.9459	0.9607	0.14

as presented in Table I and Figures 5 and 4. The analysis was based on classical comparison metrics using true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Isolation forest has been selected by AD xApp developers to be used in O-RAN software to predict the handover and optimize the wireless communication system.
However, as detailed in Table I, the results show that random forest outperforms all other ML techniques as detailed in the following:

- **Accuracy:** refers to the ratio of observations that have been correctly predicted over the total observations.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{7}$$

In our study, we found that ML techniques are pretty close in terms of accuracy ranging from 81% to 91%, followed by 74% of LOF, with the exception of logistic regression that shows a very low ratio (13%). Random Forest, however, shows a very high ratio of accuracy (98%).

- **Precision:** is the ratio of correctly predicted positive observations to the total predicted positive observations.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

By measuring the precision metric, we found out that data reveal a certain dispersion, ranging from 13% (logistic regression) to 95% (robust covariance). Random forest is on the top of the list with a precision ratio of 97%.

- **Recall:** represents the ratio of positive observations that have been correctly predicted over all observations in actual class.

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

By computing the recall metric on our ML algorithms, it appears that both algorithms, isolation forest and random forest, are the best performing with 87% and 94% respectively.

- **F-score:** is the weighted average of precision and recall.

$$F1score = 2 \cdot \frac{Recall \cdot Precision}{Recall + Precision} \tag{10}$$

The F-score results show the same performance as recall metrics, with the best performance of isolation forest (83%) and random forest (96%).

- **Root mean square error (RMSE):** is used for evaluating the prediction quality, and is based on Euclidean distance to show how far predictions fall from real values, where y_i is the actual value and $\hat{y}_i$ is the predicted value.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} ||(y_i - \hat{y}_i)||}{N}} \tag{11}$$

The RMSE has the smallest value when using random forest, which means the model is good at predicting the data, at the opposite of logistic regression where RMSE is large, meaning the model is failing to consider important features of data. In terms of training time, Naive Bayes algorithm is the fastest one spending around 38 ms in training (see Figure 4). In addition, by comparing the prediction of IF and RF performance with real data related to the normal and abnormal categories, we find that RF prediction is close to the reality as illustrated in figure 6. Given the characteristics of our dataset, RF is suitable due to the relatively low complexity of the data. However, for significantly larger datasets, the effectiveness of RF may vary. In such scenarios, it may be worth exploring a combination of different techniques such as unsupervised ML and RL techniques to detect anomalies and trigger the handover in a more efficient and effective way.

V. CONCLUSION

This paper provides an overview of the newly introduced Open RAN model that offers flexible, agile and scalable solutions

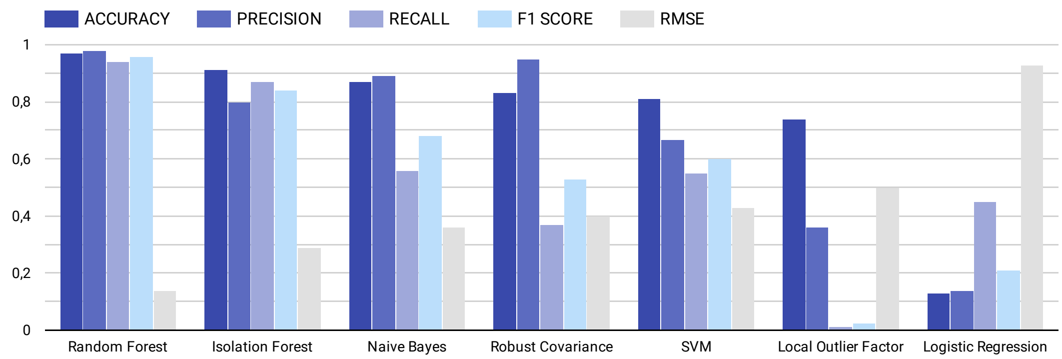

Fig. 5. Score comparison of different machine learning techniques.

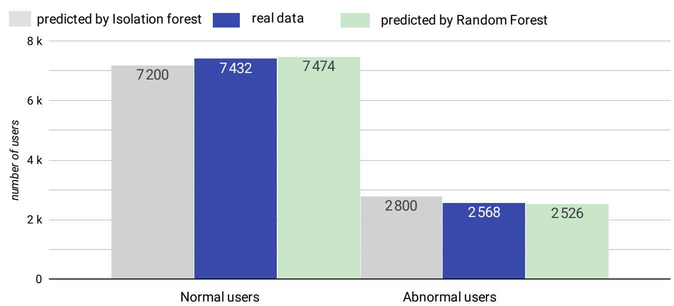

Fig. 6. Predicted performance of isolation forest and random forest.

to heterogenous wireless communication network, and to the processing of large volumes of data from diverse connected devices and users in real-time. The interoperability between RAN components and the openness of interfaces promotes the use of artificial intelligence in wireless networks, and help in optimizing services such as resource reallocation to predict congestion, smooth and seamless handover management to maintain the required QoS. Furthermore, O-RAN is designed to be able to support a large number of connected devices and users, unlike typical RAN that can not satisfy high scalability requirements. The contribution of this paper is the investigation of handover management in O-RAN, we used the traffic steering module that incorporates anomaly detection and we evaluated and compared the performance of various ML algorithms for handover prediction, including local outlier factor, isolation forest, robust covariance, logistic regression, naive bayes, SVM and random forest. Given the structure of our database, our results demonstrate that the random forest used for anomaly detection in the traffic steering module of O-RAN can significantly improve handover performance and outperform other methods. For significantly larger datasets, the classification labeling process of anomalies might be challenging and the effectiveness of RF may vary, which opens the perspectives for alternatives AI algorithms such as RL (Q-learning) or a combination of ML and RL techniques to ensure a real-time, effective amd optimized handover.

References

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE network*, vol. 34, no. 3, pp. 134–142, 2019.

[2] C. Chaccour, W. Saad, M. Debbah, Z. Han, and H. V. Poor, "Less data, more knowledge: Building next generation semantic communication networks," *arXiv preprint arXiv:2211.14343*, 2022.

[3] M. Polese, R. Jana, V. Kounev, K. Zhang, S. Deb, and M. Zorzi, "Machine learning at the edge: A data-driven architecture with applications to 5g cellular networks," *IEEE Transactions on Mobile Computing*, 2020.

[4] O-RAN, "Oran-ric-app-ad." https://github.com/o-ran-sc/ric-app-ad, 2021.

[5] L. Gavrilovska, V. Rakovic, and D. Denkovski, "From cloud ran to open ran," *Wireless Personal Communications*, pp. 1–17, 2020.

[6] H. Lee, J. Cha, D. Kwon, M. Jeong, and I. Park, "Hosting ai/ml workflows on o-ran ric platform," in *2020 IEEE Globecom Workshops (GC Wkshps*, pp. 1–6, IEEE, 2020.

[7] O. Alliance, "O-ran use cases and deployment scenarios," *White Paper, Feb*, 2020.

[8] L. Bonati, S. D'Oro, M. Polese, S. Basagni, and T. Melodia, "Intelligence and learning in o-ran for data-driven NextG cellular networks," *IEEE Communications Magazine*, vol. 59, no. 10, pp. 21–27, 2021.

[9] H. Kumar, V. Sapru, and S. K. Jaisawal, "O-ran based proactive anr optimization," in *2020 IEEE Globecom Workshops (GC Wkshps*, pp. 1–4, IEEE, 2020.

[10] S. Niknam, A. Roy, H. S. Dhillon, S. Singh, R. Banerji, J. H. Reed, N. Saxena, and S. Yoon, "Intelligent o-ran for beyond 5G and 6G wireless networks," in *2022 IEEE Globecom Workshops (GC Wkshps)*, pp. 215–220, IEEE, 2022.

[11] M. Moshtaghi, C. Leckie, and S. Karunasekera, "A framework for distributed data analysis for iot," in *Internet of things*, pp. 163–180, Elsevier, 2016.

[12] Y. Yuan, J. Yang, R. Duan, I. Chih-Lin, and J. Huang, "Anomaly detection and root cause analysis enabled by artificial intelligence," in *2020 IEEE Globecom Workshops (GC Wkshps*, pp. 1–6, IEEE, 2020.

[13] M. Sugiyama, "Statistical machine learning," *Introduction to Statistical Machine Learning*, pp. 3–8, 2016.

[14] B. Shubyn and T. Maksymyuk, "Intelligent handover management in 5g mobile networks based on recurrent neural networks," in *2019 3rd International Conference on Advanced Information and Communications Technologies (AICT)*, pp. 348–351, IEEE, 2019.

[15] S. Memon and M. Maheswaran, "Using machine learning for handover optimization in vehicular fog computing," in *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, pp. 182–190, 2019.

[16] M. Lauridsen, T. Kolding, G. Pocovi, and P. Mogensen, "Reducing handover outage for autonomous vehicles with lte hybrid access," in *2018 IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE, 2018.

[17] R. Labs, "O-ran architecture and use cases." https://www.rimedolabs.com/wp-content/uploads/2021/03/O-RAN_Webinar.pdf.

[18] O.-R. Alliance, "Traffic steering xapp and data design – bronze release." https://wiki.o-ran-sc.org/display/RICA/Traffic+Steering+Use+Case+xApps.