This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3157802, IEEE Transactions on Vehicular Technology

A Data Driven Framework for Inter-Frequency Handover Failure Prediction and Mitigation

Marvin Manalastas, Muhammad Umar Bin Farooq, Syed Muhammad Asad Zaidi, Adnan Abu-Dayya, *Senior Member, IEEE*, and Ali Imran , *Senior Member, IEEE*

***Abstract*—With 5G already deployed, challenges related to handover exacerbate due to the dense base station deployment operating on a motley of frequencies. In this paper, we present and evaluate a novel data driven solution, to reduce the inter-frequency handover failures (HOF), hereafter referred to as TORIS (Transmit Power Tuning-based Handover Success Rate Improvement Scheme). TORIS is designed by developing and integrating two sub-solutions. First sub-solution consists of an Artificial Intelligence (AI) based model to predict inter-frequency HOFs. In this model, we achieve higher than the state-of-the-art accuracy by leveraging two innovative approaches. First, we devise a novel feature set by exploiting domain knowledge gathered from extensive drive test data analysis. Second, we exploit an extensive set of data augmentation techniques to address the class imbalance problem in training the HOF prediction model. The data augmentation techniques include Chow-Liu Bayesian Network and Generative Adversarial Network further improved by focusing the sampling only on the borderline samples. We also compare the performance of state-of-the-art AI models for predicting HOF with and without augmented data. Results show that AdaBoost yields best performance for predicting HOF. The second sub-solution of TORIS is a heuristic scheme to tune the transmit (Tx) power of serving and target cells. Unlike the state-of-the-art approaches for HOF reduction that tune cell individual offset, proposed scheme targets the main cause of HOF i.e., poor signal quality and propagation condition, by proactively varying the Tx power of the cells whenever a HOF is anticipated. Results show that TORIS outperforms the state-of-the-art HOF reduction solution and yields 40% to 75% reduction in inter-frequency handover failures.**

***Index Terms*—Handover Failure Prediction, Handover Failure Mitigation, Inter-Frequency Handover, Handover Success Rate Improvement**

I. Introduction

The demand for mobile data traffic continues to grow rapidly as the number of capacity hungry devices increases. To cater this demand, 5G and beyond (5G&B) network is expected to provide massive coverage and capacity with network densification as the front runner solution [1]. However, deploying such a huge number of base stations (BSs) of different types, operating in a wide range of frequencies, makes mobility management a daunting task for the operators. This is because the number of handovers (HOs) increase proportionally with the increase in the number of BSs, and subsequently, the chances of handover failures (HOFs) also increase. The potential increase in HOFs directly affects a key performance indicator (KPI) known as handover success rate (HOSR) which leads to the degradation of other KPIs such as latency, retainability, and, throughput which ultimately results in sub-optimal user quality of experience (QoE) [2]. Optimal HO performance is particularly vital to support the ultra-reliable low-latency communication (URLLC) use case in 5G&B [3].

The success of the HO procedure mainly depends on the signal quality (i.e., noise and interference) and propagation condition between the user equipment (UE) and the source and target BSs. During the HO process, there is a back-and-forth exchange of signaling messages from the UE to source and target BS (i.e., measurement report, handover command, handover confirmation) [4]. As these messages are transmitted via the air interface, they are prone to failure if the signal condition between the UE and BS (both the source and target), is not good enough [5]. In addition, sub-optimal tuning of the handover parameters might also lead to the poor HOSR [6]–[8]. The values of these parameters are usually set based on the gold standard (GS) recommended by the equipment vendors [9]. However, gold standard values are based on one-size-fits-all approach. These values do not consider varying network deployments scenarios, cell sizes, use cases, densities, Quality of Service (QoS) requirements, and mobility patterns. The GS-based approach, therefore, cannot yield optimal performance in all cases for current and more so for future, even more dense, cellular networks. More intelligent and adaptive method to determine optimal mobility parameters is needed to minimize HOF in face of increasing cell density.

One way to reduce the HOF and to improve HOSR is to manually tune and optimize HO related parameters, i.e., hysteresis, cell individual offset (CIO) and, time-to-trigger (TTT) through hit and trial. Additionally, another industrial practice is the tuning of parameters such as tilt, azimuth, and transmit (Tx) power, which helps alleviate the HOF issues by improving the coverage and reducing interference [9]. However, such manual and mostly hit and trial-based parameter tuning is very time-consuming as the statistics have to be observed in live network for hours or days to see if the new parameter value is better than earlier. Also, dependency on human intervention and experience based incomplete or heuristic understanding of the system behavior, makes this approach prone to errors. Some self-organizing network (SON)

M. Manalastas, M.U.B. Farooq, S.M.A. Zaidi, and A. Imran are with the AI4Networks Research Center, School of Electrical & Computer Engineering, University of Oklahoma, OK, USA (email:marvin@ou.edu, umar.farooq@ou.edu, asad@ou.edu, ali.imran@ou.edu)

A. Abu-Dayya is with the Department of Electrical Engineering, Qatar University, Doha Qatar (email: adnan@qu.edu.qa)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3157802, IEEE Transactions on Vehicular Technology

functions such as mobility robustness optimization (MRO) can automate this process to some extent by automatically adjusting the CIO values based on prior HO performance [10]. Most state-of-the-art MRO solutions effectively automate the hit and trial-based tuning as a closed loop eliminating the need for manual labor. However, the fact that these SON solutions do not leverage any in-depth modeling and analysis of the system behavior, means their performance is not guaranteed to be any superior than manual hit and trial-based parameter tuning.

A. Related Work

Several studies have been done on handover failure prediction. Authors in [11] proposed a neural network capable of predicting the future Reference Signal Received Power (RSRP) values of the source and target BSs operating in similar frequency. They cascaded the RSRP prediction model with another neural network which acts as a classifier to determine if the handover will fail or not. Meanwhile, authors in [12] devised a new handover algorithm built from RSRP prediction of two remote radio units using an improved grey verhulst model. The authors proposed to do the handover decision at early step based on the predicted RSRP values. A study on forecasting future handovers is performed in [13] based on fuzzy forecasting model where authors used historical RSRP data to forecast future RSRP. While results are promising, the effectiveness of [11]–[13] heavily relies on RSRP prediction accuracy using historical data. However, signal condition such as RSRP experiences frequent fluctuations making its behavior difficult to predict and capture [14]. This intrinsic randomness in RSRP values makes RSRP prediction-based approach in predicting HOF precarious.

Literature on improving handover performance through HOF mitigation can be broadly grouped into two themes. The first theme involves utilizing the current HO standard procedure while optimizing HO-related parameters [15]–[20]. The second theme involves proposal of completely new HO algorithms [21]–[24].

Authors in [15] presented a fuzzy logic controller that modifies HO margin (HOM) to optimize call dropping ratio caused by HOFs. Meanwhile, optimization of HO parameters such as TTT and CIO to improve HOF and ping-pong has been proposed in [16]. The authors clustered users depending on the RSRP trend when users move from indoor to outdoor location. Authors in [17] used reinforcement learning to optimize parameters such as handover threshold, CIO, hysteresis, and TTT. Almost similar sets of parameters (HOM, CIO, and TTT) are analyzed and optimized in [18] with the addition of inter-site distance. Meanwhile, a single parameter, HOM, is optimized in [19] to improve HOSR taking into account the speed of the users. Authors in [20] tuned CIO for each problematic cell-pair and showed improvement in handover success rates on a real network setting. Although the results are encouraging, one caveat in HOSR improvement via optimizing HO-related parameters is that it only resolves HOFs caused by untimely handovers (i.e., too late or too early handovers). This approach fails to take into account another, and rather the main reason for HOFs and thus poor HOSR i.e., the poor signal quality and undesirable propagation condition [5].

A new HO scheme is presented [21] using a two-level neural network which predicts how user experience is affected when HO is performed to a particular BS. In [22], authors proposed a pre-handover algorithm aided by mobility prediction using Gauss-Markov mobility model to improve HOSR in Long Term Evolution (LTE). Authors in [23] proposed a new HO scheme that minimizes handover failures by reducing the HO interruption time. Another new HO algorithm is devised in [24] to improve the handover performance of users in femtocells. Authors in [24] used a completely different set of parameters compared to currently standardized set of parameters involved in HO. These include received signal strength, user speed, traffic type, and bandwidth. Although the results are promising, the proposals made in [21]–[24] are complex and require a change in the current HO standards.

Studies proposing a complete solution to mitigate HOF by leveraging a HOF prediction model are scarce. The idea is first mentioned in [11]. However, in contrast to our proposed solution that builds a HOF prediction model and uses it for enabling a HOF mitigation scheme, the study in [11] only presents a HOF prediction model and does not present a scheme to mitigate HOFs. Meanwhile, closest to our presented solutions is the work on optimization of MRO SON function in [25] where prediction of HOF is studied for improving the HOSR. However, proposed scheme uses a simple threshold-based approach for predicting HOF. It does not leverage the power of AI to predict HOF with potentially better accuracy.

Lastly, most of the current handover prediction and mitigation approaches in the existing literature are focused on intra-frequency handovers. However, in contrast to intra-frequency where both the source and target BS are on one frequency band, in inter-frequency source and target base stations have different frequencies. In inter-frequency HO the user can measure the RSRP from the target cell without switching to new frequency. For inter-frequency HO user has to periodically scan (at pre-set intervals) for other frequencies to determine potential target cells for HO. The involvement of higher number of frequency layers in inter-frequency HO necessitates the use of more input parameters into the HOF prediction model compared to intra-frequency HOF prediction model. This makes the current intra-frequency HOF prediction models ineffective for the inter-frequency HOF prediction. Additionally, to the best of author's knowledge, there is no study in the literature that aims at mitigating inter-frequency HOF. The importance of analyzing and improving inter-frequency HO further increases with the advent of 5G&B that brings with it multiplicity of frequency layers.

B. Contributions

To address the limitations of the current solutions in literature, we present a first of its kind framework, named TORIS (Transmit Power Tuning-based HOSR Improvement Scheme). TORIS consist of two components, an AI-based HOF prediction module and Tx power tuning scheme, that work in tandem to substantially reduce the HOF rate, compared to state-of-the-art schemes. The main contributions of this work can be summarized as follows:

1) We present a novel framework, named TORIS (Transmit Power Tuning-based HOSR Improvement Scheme). To the best of the authors knowledge, this is the first study to present such solution as current studies analyze HOF prediction and HOF mitigation in silos. To improve the accuracy of the TORIS's HOF prediction model we identify a novel set of features by leveraging domain knowledge and insights gained from Sobol sensitivity analysis of the raw data. This novel feature set includes features like signal strength in the form of RSRP, signal quality in the form of geometric factor (G-Factor) and user speed. These smart features improve the performance of the HOF prediction model compared to the existing models that rely only on the RSRP of the serving and target BSs [11].
2) A major challenge in HOF prediction model design is the imbalance in the real network-based training data where the number of HOF instances are far less compared to HOS instances in the gathered data. We investigate an extensive set of data augmentation techniques to address the class imbalance problem by preprocessing and augmenting the training data. We exploit an unconventional AI-based data generation technique in the form of GAN. To the best of the authors knowledge, this is the first attempt to leverage GAN in addressing the class imbalance challenge in context of cellular network. We analyze the quality of the augmented data generated by the data balancing techniques using diverse metrics. These metrics reveal that, despite their popularity, the inherent dependency of deep learning-based methods, such as GAN, on large training data make them unsuitable to generate synthetic minority data for this particular case. To further improve the performance, we leverage the borderline concept for addressing the class imbalance challenge.
3) We investigate the potential of select AI techniques for creating HOF model that provide trade-off between complexity and accuracy using the augmented as well as raw data. Results show that AdaBoost yields the best performance and thus should be a candidate of choice for creating the HOF prediction model component of TORIS in real networks.
4) We present a heuristic scheme for mitigating HOFs. This scheme constitutes the second component of TORIS and uses prediction of HOF failure model to dynamically adjust the Tx power of serving and target BSs. Results based on 3GPP compliant simulator show that the proposed HOF mitigation scheme can reduce HOF by 75%. It also leads to substantially better performance compared to the state-of-the-art CIO based schemes in literature [20].

The rest of the paper is organized as follows: Section II discusses handover success rate including the factors affecting the handover performance, simulation setup and data generation. In Section IV, we present the data augmentation techniques to address the class imbalance issue as well as the evaluation metrics to check the synthetic data quality. Section V presents the HOF prediction models and the domain knowledge-based feature engineering approach. Meanwhile, Section VI presents the HOF mitigation algorithm and its evaluation followed by the results and discussion in Section VII. Finally, Section VIII concludes the paper.

II. System Model

This section first describes the end-to-end TORIS framework for improving the HOSR by leveraging HOF prediction. It is followed by the discussion of inter-frequency HOSR and the factors affecting the handover performance.

A. *TORIS Framework*

Unlike other methods which rely on historical RSRP data, TORIS's AI-based HOF prediction module leverages measurements which are actually reported by the user to predict potential HOF occurrences. By doing this, TORIS eliminates the potential error from frequent fluctuations of radio signal condition. The advantages of the proposed Tx power tuning scheme stem from its two key attributes: 1) Compared to the state-of-the-art HOF mitigation schemes that tweak CIO, the proposed scheme addresses one root cause of HOF i.e., poor signal quality by tuning Tx powers instead of CIOs. 2) Proposed scheme is simple and does not require any change in standard and thus can be easily implemented in real 4G, 5G and beyond networks. These advantages stem from the fact that while 5G&B network has several new features (i.e., mmWave utilization and network slicing) vis-à-vis legacy networks such as 4G, 5G still leverages the same handover standard procedure as 4G network [26]. Therefore, irrespective of the adaptation of new physical layer technology such as mmWave or system level orchestration such as network slicing, the proposed solution would work for 5G since it is designed based on the 5G standard handover procedure.

TORIS framework is illustrated in Fig. 1. TORIS starts with data generation and collection from sources like minimization of drive test (MDT), drive test and Operations Support Systems (OSS) data as depicted in the top left block in Fig. 1. In this paper, data generation is done using a 3GPP-compliant system level simulator. As in a typical commercial real network, HOF events happen far less frequently compared to handover success events, the training data that can be harnessed to train a HOF prediction model is bound to be extremely imbalanced. This imbalanced data, if used as it is for training an AI model, can lead to a model biased towards misclassifying all or most HO events as HO success events. To address this challenge, TORIS leverages several data augmentation and synthetic data generation techniques as depicted by the TORIS modules in 2nd row and middle column of Fig. 1. These techniques include Synthetic Minority Oversampling TEchnique (SMOTE), Chow-Liu Bayesian Networks (CLBN), and Generative Adversarial Network (GAN) based deep neural networks combined with border line sampling concept. However, the data generated using these techniques are not readily usable. By incorporating a data evaluation process (top

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3157802, IEEE Transactions on Vehicular Technology

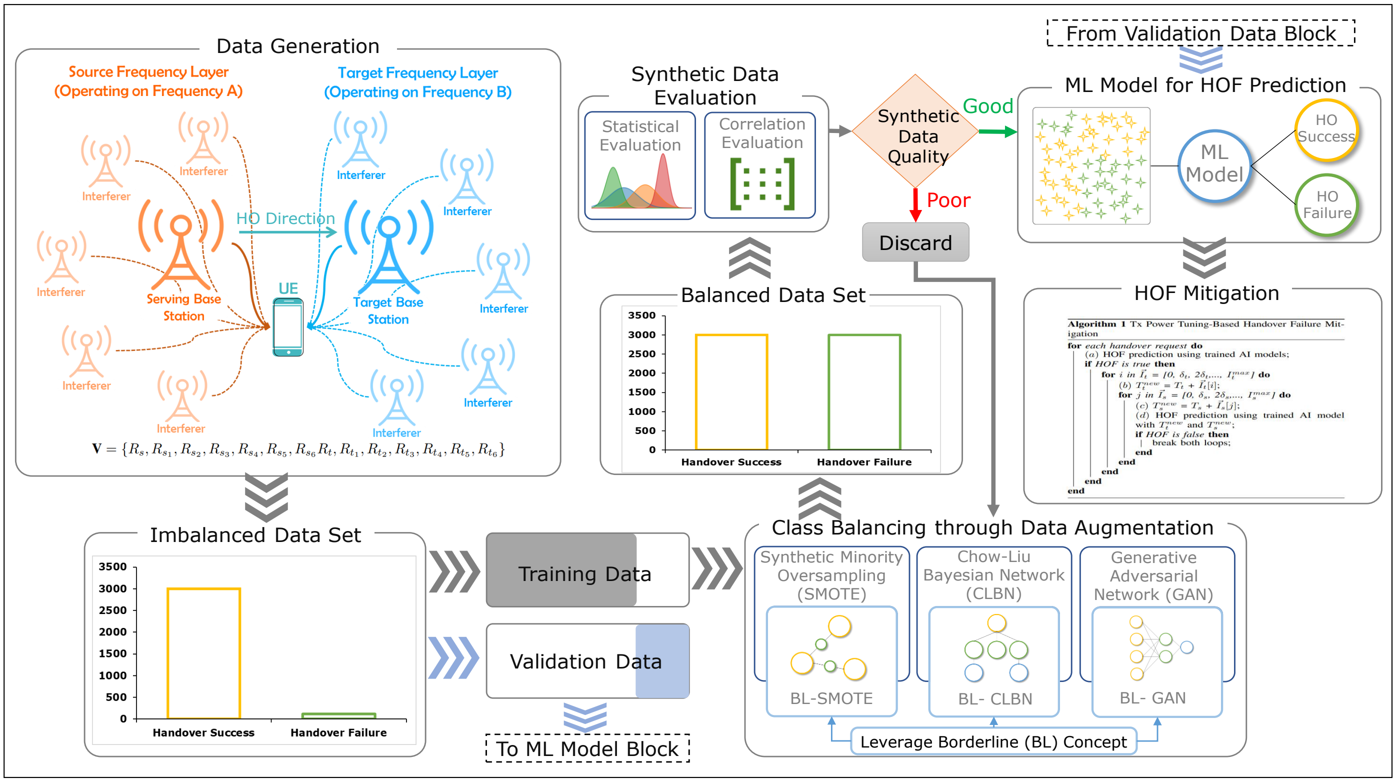


Fig. 1. TORIS (Transmit Power Tuning-based HOSR Improvement Scheme) framework.

right block in Fig. 1), TORIS ensures that the quality of the generated data is acceptable. Synthetic data with good quality are used to train different ML models providing a multi-faceted performance comparison of Naïve Bayes Classifier, K-nearest Neighbors (KNN), Support Vector Classifier, Random Forest, Extreme Gradient Boosting (XGBoost), and Adaptive Boosting (AdaBoost) for predicting HOFs. These state-of-the-art methods are selected based on their effectiveness as binary classifiers and their ability to work well specially on small data sets compared to deep neural networks, in addition to the faster training which is crucial for HOF prediction. Further analysis of the advantages of the selected six techniques for binary classification, compared to other ML methods, reader is referred to [27]–[29]. Whenever ML detects potential HOF, the HOF mitigation scheme kicks in (3rd block from top right) and rectify the failure by heuristically varying the transmit power of the involved BSs.

B. Inter-frequency Handover Success Rate

In current 4G and 5G networks, there are two kinds of handovers namely intra-frequency handover and inter-frequency handover. The former happens when a user switches from one BS to another operating on the same frequency band. The latter happens between the cells operating on different frequencies. Inter-frequency handovers also happen when the UE requests a service which is not available at current frequency or during load balancing across the frequency bands. Inter-frequency HO utility is expected to expand due to the co-existence of 4G and 5G at different bands. This signifies the importance of inter-frequency HOSR optimization for the current and future cellular networks. However, despite its growing significance and associated open challenges, inter-frequency HOSR enhancement remains an under investigated topic as most literature remains focused on intra-frequency handover.

Inter-frequency HOSR is a direct measure of the handover performance between different frequency bands. For this reason, HOSR is one of the major KPI of interest in the current and emerging cellular networks. High HOSR means that the movement of users from one BS to another is seamless and thus, satisfactory QoE is maintained. In addition, URLLC, which is one of the use cases of 5G and beyond heavily relies on seamless handovers particularly for applications such as intelligent transport systems and autonomous cars. HOSR ξ can be described as:

$$\xi = \frac{n_s}{n_s + n_f} \times 100\%, \tag{1}$$

where n_s and n_f are the number of successful and failed inter-frequency handovers, respectively, in the network.

C. Factors Affecting Handover Success Rate

Aside from sub-optimally configured handover parameters [15]–[20], one of the major factors affecting HOSR is signal strength and quality of both the source and the target BSs. As previously mentioned, handovers are executed with the help of interchanging commands and acknowledgments between the UE and the BS. During this crucial period, the connection between the UE and the BS should be good enough to maintain the back-and-forth exchange of HO-related messages. In current cellular network, signal strength is measured by the parameter called RSRP. The downlink RSRP R_u^x for a user u connected to the serving BS trying to perform HO to the target BS is given by:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3157802, IEEE Transactions on Vehicular Technology

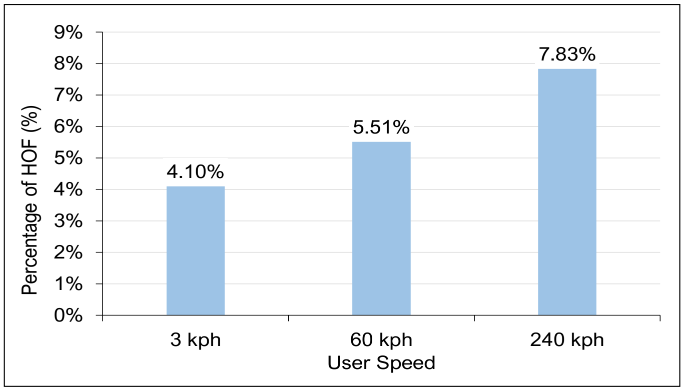


Fig. 2. Impact of user speed in the occurrence of HOFs.

$$R_u^x = T_x d_u^x, \tag{2}$$

where x corresponds to either serving or target BSs, T is the transmit power of the BS and d is the path loss dependent component of the user u. The path loss dependent component also contains antenna gains as well as the shadowing for the user, which is modeled as a Gaussian random variable.

Meanwhile, the quality of the signal in a cellular network can be measured using several metrics. In this paper, we use signal quality metric known as G-Factor. G-Factor can be defined as the ratio between the RSRP and the combined signal power of interfering cells in downlink. Mathematically, G-Factor is expressed as:

$$G_u^x = \frac{R_u^x}{\sum\limits_{\forall i \in I_x} R_u^i + n_0}, \tag{3}$$

where G is the G-Factor, x is either the serving or target BS, the set I_x contains 5 strongest interferers for the BS x and n_0 is the noise floor constant. The selection of 5 strongest interferers is based on the industrial domain knowledge of the authors. Based on our field experience of working with real 4G and 5G networks, the interference from 5 neighboring base stations contributes the most in the calculation of G-Factor and the impact diminishes from sixth interferer onwards.

Although signal-to-interference-plus-noise ratio (SINR) is also a good metric to measure the signal quality, it is sensitive to instantaneously changing load of the neighboring BSs. To eliminate this dependency, we use G-Factor instead. In addition, as G-Factor is basically the worst case SINR, using it as a metric to predict the HOF will cover worst case instances which might be missed if SINR is used as signal quality metric.

In addition to the signal strength and quality, we have also used user speed as an input feature for HOF prediction. Speed is shown to have an impact on handover performance [30]. This study showed a decrease in handover success rate with the increase in user speed. Our own analysis on the effect of user speed confirms the finding in [30]. Fig. 2 shows the increase in the percentage of HOFs with increase in the speed of the users.

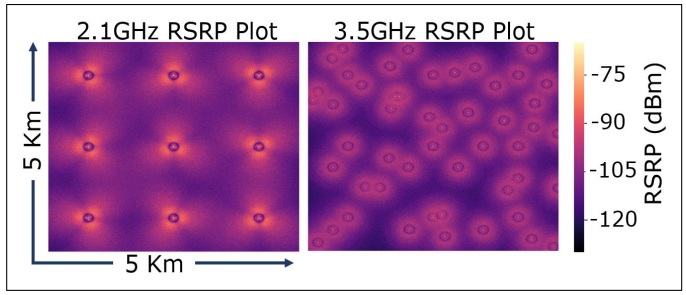


Fig. 3. Network deployment layout and RSRP plot for simulations.

TABLE I
DESCRIPTION OF SIMULATION PARAMETERS

Parameter Description	Value
Simulation area	25 km^2
Number of 2.1GHz macro transmitters	27
Number of 3.5GHz small cells	50
Macro cell height	30 m
Small cell height	20 m
Macro cell Tx power	40 dBm
Small cell Tx power	30 dBm
Pathloss exponent	3
Shadowing standard deviation	4
Active user density λ_u	12 per km^2
Speed vector S	[3, 60, 240] km/h
Transmission time interval (TTI)	1 ms
Total simulation time	15000 ms

III. Simulation Setup and Data Generation

Collecting the necessary data from a live network though plausible in theory, is impractical in practice. Due to the sporadic nature of HOF occurrence, numerous drive tests are necessary to gather enough samples. This process of gathering data samples is an extremely resource-intensive task. In this backdrop, to generate the data, we exploit a 3GPP-compliant state-of-the-art system level simulator named SyntheticNET [31]. This is the first simulator to model 5G mobility parameters in detail as needed for this study. SyntheticNET simulator has been calibrated against real network measurements to ensure the validity of the data generated through it.

An area of size 5km x 5km is used for the simulation as shown in Fig. 3. A multi-carrier heterogenous network composed of two frequency layers is deployed inside the area. Macrocells operating at 2.1GHz, with 3 sectors each, are deployed using grid pattern. Meanwhile, small cells operating at 3.5GHz, with omni-directional antenna, are deployed in the network following a uniform random distribution. The initial deployment of the users in the network follows a uniform distribution with user density λ_u. Each user can move in the network with speed s_u chosen from a set S. All elements of the set S are equally probable, and the speed value remains constant for a user. The user mobility type is a random way point. The network level simulation parameters are summarized in Table I.

The initial mobility parameters are chosen based on the GS setting of one of the leading operators in the USA. Event A2, which is triggered when RSRP of the serving BS becomes lower than a threshold, is used to trigger the measurement gap as a pre-requisite for inter-frequency handover. In the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3157802, IEEE Transactions on Vehicular Technology

simulations, we used -100 dBm for A2-threshold. Meanwhile, to trigger the HO process, event A3 is used. This event is triggered when the RSRP of the target BS becomes better than the serving BS's RSRP by some offset value. A3-offset of 2dB is used. TTT of 16ms and hysteresis of 0dB was used for events A2 and A3.

To gather data for handover failure prediction, we run the simulation for around 15000 TTIs. While moving, the users inside the simulation area measure and log the RSRP condition of the serving as well as the neighboring base stations in both frequency bands in the form of a state vector ($\mathbf{V}$) which can be expressed as:

$$\mathbf{V} = \{R_s, R_{s,n1}, R_{s,n2}, R_{s,n3}, R_{s,n4}, R_{s,n5}, \\ R_t, R_{t,n1}, R_{t,n2}, R_{t,n3}, R_{t,n4}, R_{t,n5}\}, \quad (4)$$

where R is the reported RSRP, subscripts s and t correspond to the serving and target BSs operating on different frequency bands, respectively and $n_i \forall \; i \in 1,2,3,4,5$ relate to the interfering neighbors in the serving and target frequency layer sorted in descending order. When handover condition is triggered, the base station starts the HO process. After the HO execution, the measured RSRPs are flagged with either handover success or handover failure depending on the outcome of the handover. The resulting combination of RSRPs and handover status (i.e., failed or successful) are used to train HOF prediction models (after going through the data augmentation process to address the class imbalance issue). After the HOF prediction model is trained offline, it is imported into the handover failure algorithm to make online inference. A new simulation with similar setting as the simulation done for HOF prediction data generation is run with the addition of the HO failure mitigation algorithm and handover performance is observed. It is worth highlighting that similar simulation setup is used for objective evaluation of the gain of the proposed scheme. The developed HOF prediction model is expected to work in varying deployment scenarios as it relies on user speed, RSRP, and G-factor of both the serving and target BS, and not on the actual deployment scenario or other configuration parameters in the simulation setup.

IV. Data Augmentation and Evaluation of Synthetic Data Quality

In this section, we analyze the class imbalance issue and solve this problem using several synthetic data augmentation techniques. As a form of sanity check for the generated synthetic data to address the class imbalance issue, we present methods of evaluating the synthetic data quality.

A. Effect of Imbalance Data in HOF Prediction

Although handover failures are common, the occurrences are far lesser than the number of successful handovers. Since the simulator we used models a realistic HO scenario, our data set follows a similar trend. Majority of the samples are HOSs with very few HOFs. A classification model trained on the imbalanced data set might lead to the model poorly learning

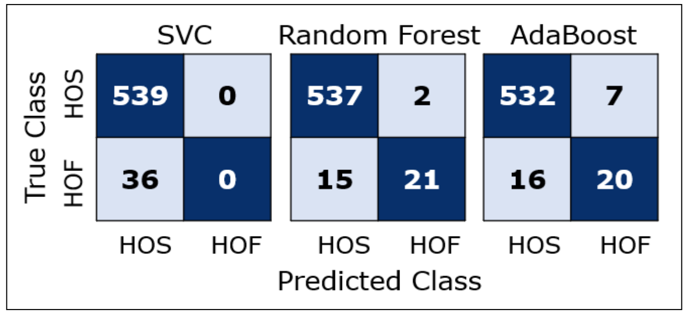


Fig. 4. Confusion matrix of the classifiers trained using imbalanced data set.

the decision boundary. The classification model becomes biased towards the majority class, making the prediction in favor of the majority class (i.e., HOS) while neglecting the more important minority class (i.e., HOF). Fig. 4 shows the impact of the imbalanced data set on the accuracy of some state-of-the-art classifiers. Worst classifier, Support Vector Classifier in this case, fails to predict even a single instance of HOF correctly. Meanwhile, two best classifiers Random Forest and AdaBoost are able to correctly classify only 21/36 (58.3%) and 20/36 (55.5%) HOFs, respectively.

B. Data Augmentation Techniques to Address Class Imbalance

The class imbalance issue in classification can be solved by oversampling the minority class. Generating more samples of the minority class can improve the decision boundaries by making them less specific to the majority class. This wider decision boundaries improve the classifier generalization of the data set which ultimately improves the classification performance. In this paper, we exploit three types of synthetic data generation techniques. First, we leverage the most commonly utilized technique in synthetic data generation for imbalanced data set called SMOTE, originally proposed in [32]. We use SMOTE for its well-known ease of implementation and interpretability. Second, Bayesian Network (BN) is another type of synthetic data generation technique, which creates an acyclic graph using probabilistic models. We choose BN for this study as it is simple, fast, and known to work efficiently even with small data samples [33]. We use Chow-Liu [34] method that approximates the tree with first-order dependency, which has the smallest Kullback–Leibler (KL) divergence to the actual joint probability distribution. In this paper, we utilized the algorithm given in [35] for CLBN-based data generation. Finally, we use some newer synthetic data generation methods in the form of GAN. We utilize GAN due to its recent promising results in augmenting tabular numerical data. In this paper, we leverage conditional tabular GAN (CTGAN) recently proposed in [36] to generate synthetic data. This particular type of GAN has been shown to outperform other variation of GANs for tabular data.

For optimal performance, we tune and evaluate several hyperparameter combinations of CTGAN as shown in Table II. Based on this list of hyperparameters, we perform grid search to find the best hyperparameter combination. We perform the evaluation by generating synthetic data using the different settings of CTGAN and training the AI-model to predict HO failures. We select the hyperparameter combination that produces

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3157802, IEEE Transactions on Vehicular Technology

TABLE II
HYPERPARAMETERS USED FOR TRAINING CTGAN

Parameter Name	Values Evaluated	Optimal Setting
(Embedding Dimension) - (G Dimension) - (D Dimension)	(128) - (256,256) - (256,256) (64) - (128, 128) - (128,128) (32) - (64, 64) - (64, 64) (16) - (32, 32) - (32, 32) (8) - (16, 16) - (16, 16) (4) - (8, 8) - (8, 8)	(32) - (64, 64) - (64, 64)
Adam Optimizer Weight Decay	1e-01, 1e-03, 1e-05, 1e-07	1e-07
Batch Size	10, 30, 50, 70	50
Epoch	100, 300, 500, 700	500

the highest area under receiver operating characteristic (AUROC) curve and area under precision recall curve (AUPRC). The best performing CTGAN hyperparameters shown in Table II are selected and the performance in predicting HO failures is compared with other data augmentation techniques.

C. Data Augmentation Leveraging the Borderline Concept

Since its debut, several variations of SMOTE have been devised to improve its performance. One such variation is called Borderline-SMOTE (BL-SMOTE) introduced in [37]. BL-SMOTE has been proven to improve the performance of standalone SMOTE based on studies conducted by authors in [38], [39]. BL-SMOTE works on the principle of SMOTE with one major modification. It uses KNN to identify the samples near the decision boundary and over sample them instead of blindly oversampling all the minority data points. This improves the performance of SMOTE by limiting its action on samples which actually require oversampling.

Inspired by this concept, we implement the borderline technique with CLBN and GAN which we call BL-CLBN and BL-GAN, respectively. Just like BL-SMOTE, an additional step is implemented that first determines the minority samples near the class boundary using KNN. These samples are then fed into the CLBN and GAN to generate synthetic data. This method provides more resolution on the important area to facilitate better performance than simple CLBN and GAN.

D. Evaluation Methods for Synthetic Data

A thorough evaluation of the synthetic data generated from the aforementioned techniques is required before using it for model training because synthetic data, in some cases, are not accurate representation of the actual data. To ensure the synthetic data represent the actual data distribution, we conduct two types of evaluations to examine the quality of the generated data described below.

1) Statistical Evaluation: Earth Movers Distance (EMD) also known as Wasserstein Metric is used to measure the similarity between the distribution of real and synthetic data. The lower the EMD is, the closer one distribution is to the other. EMD is considered more robust compared to other distance metrics such as KL divergence and Jensen–Shannon divergence especially when the two distributions are disjoint or not in the same probability space [40]. Mathematically, weight transferred EMD (ζ) can be expressed as:

$$\zeta = \sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{W}_{ij}d_{ij}, \tag{5}$$

where $\mathbf{W}$ is a matrix and each element corresponds to the amount of mass transferred from one distribution in the ith location to the other distribution in the jth location while d is a distance affinity matrix which corresponds to the cost of moving a single element of mass from the ith location of one distribution towards the jth location of the other distribution. This distance is calculated as the Euclidean ground distance between the two distributions.

2) Pairwise Correlation Comparison: Although EMD can capture the deviation of the distributions between real and synthetic data, it cannot capture the changes in the relationship between each input variable. To address this problem, we use pairwise correlation difference (PCD) as an additional evaluation method of synthetic data. The PCD metric measures the difference in the Frobennius norm ($\|\ \|_F$) of the Pearson correlation ($Corr$) between real and synthetic data sets [41]. Mathematically, PCD is defined as:

$$PCD(M_{real}, M_{synt}) = \|\ Corr(\mathbf{M}_{real}) - Corr(\mathbf{M}_{synt})\ \|_F, \tag{6}$$

where $\mathbf{M}_{real}$ and $\mathbf{M}_{synt}$ are the real and synthetic data matrices, respectively. Good synthetic data should have close to zero entries inside PCD matrix. Additionally, to represent PCD using a single variable instead of a matrix, we use the variable μ which corresponds to the mean absolute value of the elements in the PCD matrix. Mathematically, μ is defined as:

$$\mu = \frac{1}{n}\sum_{c=1}^{C}\sum_{r=1}^{R}|PCD(M_{real}, M_{synt})_{rc}|, \tag{7}$$

where $|PCD(M_{real}, M_{synt})_{rc}|$ is the absolute value of an element from the PCD matrix in rows r and columns c while R and C are the number of rows and columns of the PCD matrix, respectively. Meanwhile, n is the total number of matrix elements.

V. AI-BASED HANDOVER FAILURE PREDICTION MODEL

In this section, we discuss the HOF prediction approach and the AI models we leverage in this paper. We also present the domain knowledge aware feature engineering to improve the prediction performance against state-of-the-art solution.

A. HOF Prediction Models

We model HOF prediction as a binary classification problem. With classification, labels of the data points are predicted by mapping input features $\mathbf{V}_o$ to discrete labels y. In our study, input features $\mathbf{V}_o$ presented in eq. (8) include the measurements that a UE sends to the BS which trigger handover. These include the RSRP values of the serving and target BS as well as up to 5 strongest interferers for both serving and target layer. UE can report these RSRP values when the event is triggered. The RSRP values are saved at the start of the TTT when the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3157802, IEEE Transactions on Vehicular Technology

entering condition of the event becomes true for the first time. Each of the input combinations $\mathbf{V}_o$ are labeled as either HOS or HOF. This HOS and HOF represents the discrete class y. Since the simulator is already calibrated against real network, the general insights drawn from the model are expected to hold for a real network. Nevertheless, for practical use of this solution, new model will have to be trained for each deployment scenario using real data where possible.

In this paper, we evaluate the performance of six classification techniques namely Support Vector Classifier, Naïve Bayes, KNN, Random Forest, XGBoost, and AdaBoost. These state-of-the-art methods are selected based on their effectiveness as binary classifiers and their ability to work well specially on small data sets compared to deep neural networks, in addition to the faster training which is crucial for HOF prediction. Below are the discussions of each of the ML model we leverage:

1) Support Vector Classifier (SVC): SVC is a type of supervised ML technique which is often used for binary classification problems. Basically, SVC uses a hyperplane which is designed to classify training vectors into two classes. Its popularity can be attributed to its simplicity, speed, and effectiveness to work very well with limited amount of data points. However, SVM works best on data sets which are linearly separable.

2) Naïve Bayes Classifier: As the name implies, Naïve Bayes classifier works based on Bayes theorem. This ML algorithm assumes independence among the input features. That is, each feature in a class in unrelated to other features making each of them contribute independently to the classification action. Naïve Bayes classifiers are simple, easy to build and work best for large data sets.

3) k-Nearest Neighbors (KNN): KNN is one of the simplest ML algorithms and uses the concept of similarity measure such as distance function to make classification decisions. It stores all the available cases and classifies a new data point based on the class of neighbors.

4) Random Forest: Random Forest is another kind of tree-based learning algorithm composed of sets of decision trees. These trees are built using randomly sampled subset of the whole training data. The final class is decided by majority voting from the class results of all decision trees. Random Forest is considered as one of the most reliable algorithms for classification problems due to high accuracy even with small data set and large input features.

5) Extreme Gradient Boosting (XGBoost): XGBoost is a popular type of gradient boosting algorithm which belongs to the ensemble learning category of ML. Ensemble learning techniques train several learners to perform the same task. XGBoost trains multiple regression trees called weak learners. These weak learners are then converted into a single superior learner to combine the decision results of all the weak learners.

6) Adaptive Boosting (AdaBoost): AdaBoost is another type of ensemble learning. It is composed of several weak decision trees with single split called decision stumps. For improved performance, AdaBoost puts more weight on difficult classification instances while giving less weight on easy to classify samples.

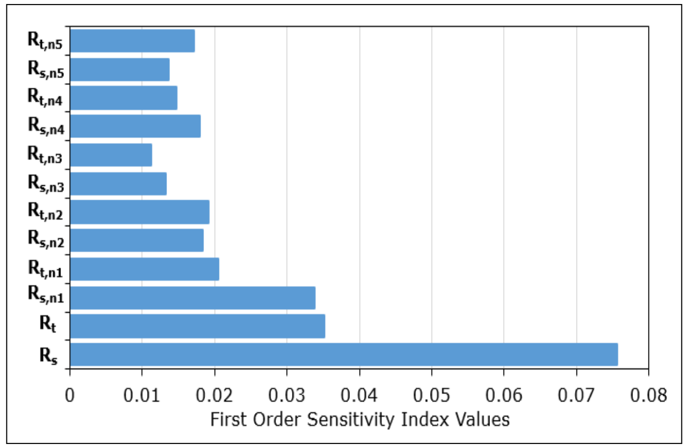


Fig. 5. Sobol sensitivity analysis of the input features used for HOF prediction.

B. Domain Knowledge Based Feature Engineering

The performance of an AI model depends on the quality of the input features. Combined with domain aware smart selection of the input features, quality of the AI model can be improved. To this end, we begin by performing a detailed exploratory analysis to extract the features that influence the handover failures the most. As a first step, we perform the Sobol-based variance sensitivity analysis [42]. The findings of which is presented in Fig. 5. It is observed that the RSRP of the source BS has the highest impact on the performance of the prediction model followed by the RSRP of the target BS while the Top 2 interferers for the source and target are in third and fourth spots, respectively. Using these insights from the Sobol analysis, we identify sets of input features, which can potentially predict the HOF. The first feature set consists of the most important feature (source RSRP), while the second set consists of the two most important features (source and target RSRP). The feature set 2 has already been used to predict HOF in a previous study [11]. Set 4 consists of four most important features (source RSRP, target RSRP, top interferers of the source and target BS). Feature set 6 contains all the 12 input features without any intelligent feature selection. We leverage domain knowledge to incorporate the impact of all interferers using a single parameter in feature set 8. To evaluate the impact of user speed on each of the aforementioned feature sets, we devise sets 3, 5, 7 and 9.

A comparison of the feature sets intelligently derived using Sobol analysis is given in Table III. The analysis of feature sets 1, 2, 4, 6 in terms of area under receiver operating characteristic (AUROC) curve and area under precision recall curve (AUPRC) highlights that the input feature set 4 that includes four most important features performs better than the feature sets 1, 2 and 6. The better performance of feature set 4 compared to feature set 6 shows that the HOF prediction does not improve by just adding more interferers as input features. Meanwhile, adding user speed as input features (i.e., sets 3, 5, 7) showed improvement in the AUROC and AUPRC performance compared to their counterpart sets 2, 4, and 6, respectively.

To this end, we devise feature set 8, where G-Factor is

TABLE III
DIFFERENT SET OF INPUT FEATURES AND THEIR EFFECT IN HOF PREDICTION

Set #	Feature Vector	AUROC	AUPRC
1	$\{R_s\}$	0.877	0.635
2 [11]	$\{R_s,R_t\}$	0.875	0.628
3	$\{S,R_s,R_t\}$	0.884	0.638
4	$\{R_s,R_t,R_{s,n1},R_{t,n1}\}$	0.919	0.643
5	$\{S,R_s,R_t,R_{s,n1},R_{t,n1}\}$	0.920	0.637
6	$\{R_s,R_{s,n1},R_{s,n2},R_{s,n3},R_{s,n4}, R_{s,n5}, R_t,R_{t,n1},R_{t,n2},R_{t,n3}, R_{t,n4},R_{t,n5}\}$	0.916	0.635
7	$\{S,R_s,R_{s,n1},R_{s,n2},R_{s,n3},R_{s,n4}, R_{s,n5}, R_t,R_{t,n1},R_{t,n2},R_{t,n3}, R_{t,n4},R_{t,n5}\}$	0.928	0.643
8	$\{R_s,R_t,G_s,G_t\}$	0.947	0.698
9	$\boldsymbol{\{S,R_s,R_t,G_s,G_t\}}$	**0.951**	**0.713**

used as an input feature instead of raw interference. G-factor combines the impact of all the interferers in one input feature as described in eq. (3). Results in Table III show that this domain knowledge aware smart feature improves the AUPRC of feature set 8 compared to other aforementioned feature sets. Finally, the speed of the users is also included as a new input feature to generate feature set 9. Results show that the input feature set 9 with user speed and G-Factor yields better performance compared to all other feature vectors with AUROC and AUPRC of 0.951 and 0.713, respectively.

Results from the feature engineering analysis led to the formulation of the optimal input feature vector $\mathbf{V}_o$ expressed as:

$$\mathbf{V}_o = \{S, R_s, R_t, G_s, G_t\}\,, \tag{8}$$

where S corresponds to user speed, R and G correspond to RSRP and G-Factor, respectively with subscripts s and t correspond to the serving and target inter-frequency BSs, respectively.

VI. HANDOVER FAILURE MITIGATION ALGORITHM AND PERFORMANCE EVALUATION

This section focuses on the presentation of the HOSR improvement algorithm to complete the second sub-solution of the TORIS framework. Moreover, we also discuss the method of evaluation to measure the effectiveness of the proposed algorithm.

A. Tx Power Tuning based HOSR Improvement Algorithm

The insights gained from the HOF prediction model discussed in Section V, we conclude that HOFs occur due to certain combinations of RSRP of the source and target BSs. Modification of soft parameters such as CIO is one way to improve HO performance. Handover can happen earlier or later with CIO tuning and hence changing the RSRP conditions of source and target BS during handover. However, CIO cannot be tuned for every user and can result in instances of HOFs for different set of users performing HO between the same source and target BS. To solve this problem, we propose to mitigate HOF for each type of HO instance by changing the Tx power of the serving and target BS instead of tuning soft parameters such as CIO [20].

Algorithm 1 presents the proposed method to mitigate HOF by varying the Tx power of the source and target BSs. In each handover request made by the UE, the reported measurements are fed into the trained AI classifier. For each input combinations, the AI model detects if the requested HO will fail or not. In case of a handover success prediction, the source BS commence with the handover to the target BS. However, if there is a predicted HOF due to the received measurement, the source BS incrementally increases its Tx power. Increasing the Tx power is the method of choice since decreasing the Tx power might lead to further degradation of signal condition and quality, which will lead to more chances of HO failures.

While the increase in the Tx power through TORIS framework is effective in most cases with poor signal condition at the HO point, the increment in the Tx power might have less impact in certain network scenarios. For instance, if there is a sudden decline in source RSRP due to some blockage, TORIS algorithm can cause further delay in the HO. In such cases, reducing the Tx power of the source BS might be more appropriate. Furthermore, decreasing the Tx power of the target cell can avoid wrong or unwanted HO.

The increment in source Tx power is chosen from source BS increment vector $\vec{I_s}$ = [0, δ_s, $2\delta_s$,..., I_s^{max}], where δ_s and I_s^{max} are the increment value and maximum power increase for source BS, respectively. In each iteration, the new source Tx power T_s^{new} is computed by adding a value of increment from $\vec{I_s}$ in existing source Tx power T_s. Increasing the Tx power of source BS is given preference over the target BS because it has the highest impact on HOF as shown in Fig. 5. For each increment made, the algorithm feeds back the new measurement combination to the HOF prediction model. The model then checks if the new combination results in HOF or not. Once the maximum power increment in the source BS reaches the limit and the HOF prediction model still detects unsuccessful HO, the algorithm increments the Tx power of the target BS. The increment in target Tx power is chosen from target BS increment vector $\vec{I_t}$ = [0, δ_t, $2\delta_t$,..., I_t^{max}], where δ_t and I_t^{max} are the increment value and maximum power increase for target BS. In each iteration, the new target Tx power T_t^{new} is computed by adding a value of increment from $\vec{I_t}$ in existing target Tx power T_t. The Tx power of source and target BS are increased to T_s^{new} and T_t^{new}, respectively only when the AI model indicates a successful HO. Similar to CIO values, the increment and limit for maximum change in the Tx power can be controlled by the operator. In this paper, we use 1dB value for δ_s and δ_t and 3dB value for I_s^{max} and I_t^{max}. HO process commences only after a HOS prediction by the AI model or the maximum limit for both source and target Tx power variation is reached.

Since the increment in Tx power only occurs momentarily for a specific signaling message, the probability of two or more neighboring base stations to transmit handover signaling messages at the exact time instant is very low. However, for a very rare case wherein two or more base stations transmit signaling message and increase the Tx power at the same time, the performance of the proposed algorithm might degrade due

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3157802, IEEE Transactions on Vehicular Technology

to potential momentary increase in interference and decrease in the SINR. In addition, the increase in Tx power for one TTI will impact other UEs in the network only during this TTI. The signal strength of the users connected to the base station increasing the Tx power will improve but the SINR of the users connected to neighboring base stations (operating at the same frequency) will decrease. Similarly, the momentary increase in the Tx power is not enough to trigger unnecessary handover due to its much shorter duration (1ms) compared to handover TTT (mostly more than 40ms). Therefore, the impact on other UEs in the system will be for one TTI and very sparse in time.

Algorithm 1 Tx Power Tuning-Based Handover Failure Mitigation

for *each handover request* **do**
  (a) HOF prediction using trained AI models;
  if *HOF is true* **then**
    for i *in* $\vec{I}_t = [0, \delta_t, 2\delta_t, \ldots, I_t^{max}]$ **do**
      (b) $T_t^{new} = T_t + \vec{I}_t[i]$;
      for j *in* $\vec{I}_s = [0, \delta_s, 2\delta_s, \ldots, I_s^{max}]$ **do**
        (c) $T_s^{new} = T_s + \vec{I}_s[j]$;
        (d) HOF prediction using trained AI model with T_t^{new} and T_s^{new};
        if *HOF is false* **then**
          break both loops;
        end
      end
    end
  end
end

B. Handover Failure Mitigation Scheme Evaluation

Inaccuracies in HOF detection model creates two scenarios. First, handover failure mitigation may not activate due to incorrect prediction of HOF as HOS. This shortcoming is captured by evaluating the number of uncompensated HOF. These instances refer to the cases when HOF are not predicted or resolved correctly. Second scenario happens when HOF mitigation is falsely activated due to wrong prediction of potential HOS as potential HOF. To capture this, we have measured the number of false compensations, which are the recorded number of instances wherein the algorithm kicks in to compensate a handover request unnecessarily. False compensations in addition to delay in HO can lead to waste in power resources due to the unwanted increase in Tx power. A slight instantaneous increase in interference on the particular signaling messages, to nearby users if they are going through HO at the same time with neighboring BSs, is another cost of false compensation. Since increase in Tx power is done for only certain TTIs, the impact of false compensation rate can be determined by the number of false compensations affected TTIs over the total simulation TTI (i.e., 15000ms).

We measure the performance of proposed HOF minimization scheme by defining a metric which captures both uncompensated HOF and false compensations. Firstly, the number of uncompensated HOFs (Δ) due to errors in HOF prediction can be written as:

$$\Delta = \Theta_t - \Theta_c, \tag{9}$$

where Θ_t represents the total number of HOFs and Θ_c shows the total number of compensated HOFs. Since ideally, all HOF should be resolved, the value of Δ should be as small as possible. Secondly, false compensation (Γ) also causes negative impact on the network performance and thus should be as minimal as possible. Therefore, the combined failed compensation and false mitigation score Υ is calculated as:

$$\Upsilon = (\alpha \times \Delta_{norm}) + [(1-\alpha) \times \Gamma_{norm}], \tag{10}$$

where Δ_{norm} and Γ_{norm} are the normalized values of uncompensated HOFs Δ and false compensations Γ, respectively, α corresponds to the operator-defined weight with values from 0 to 1. Operator-defined means that the network operator can choose the value of α depending on their preference. α can be used to set the priority between minimizing uncompensated HOFs and false compensations. For instance, a high value of α will give more priority to uncompensated HOF minimization compared to potential false compensation. Similarly, it can be set to 0.5 to give equal priority to uncompensated HOFs and false compensation. For different values of α, best augmentation technique can be evaluated as the one having the least value of false mitigation score Υ.

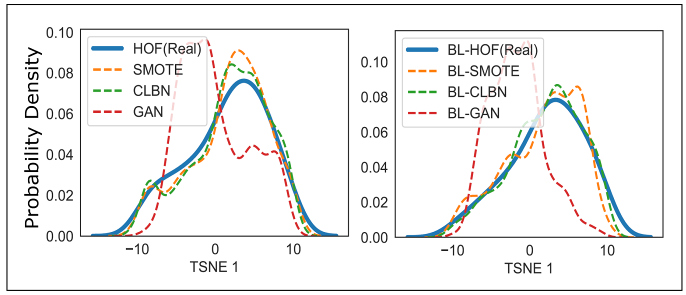


Fig. 6. PDF comparison between real and synthetic data using different data augmentation techniques.

VII. Results and Discussion

A. Synthetic Data Evaluation Results

1) Statistical Evaluation Results: Tabular synthetic data is generated by SMOTE, CLBN, GAN, BL-SMOTE, BL-CLBN and BL-GAN. Each technique produces five column vectors, and each column corresponds to an input feature similar to the real data. To overcome the complexity of comparing the real and synthetic data using each column individually, we compare them using a single parameter that captures the likelihood of the real and synthetic data as a whole. We achieved this goal through a dimensionality reduction technique called t-Distributed Stochastic Neighbor Embedding (t-SNE). This method transforms the input vector with 5 dimensions into a single t-SNE component. Fig. 6 shows the probability distribution function (PDF) of the single t-SNE component of real and synthetic data. From the results in Fig. 6, we see SMOTE and CLBN synthetic data follow the distribution of the real data very closely. However, distribution of data

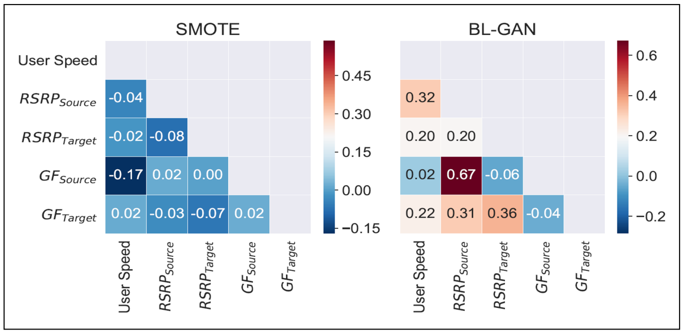


Fig. 7. Best and worst pairwise correlation difference comparison.

TABLE IV
SUMMARY OF EVALUATION COMPARISON BETWEEN DIFFERENT DATA AUGMENTATION TECHNIQUES

	EMD (ζ)	Mean PCD (μ)
SMOTE	0.325	0.047
BL-SMOTE	0.365	0.101
CLBN	0.312	0.054
BL-CLBN	0.230	0.106
GAN	2.040	0.177
BL-GAN	3.953	0.238

produced using GAN varies from the distribution of the real data. When using the borderline technique, results show that BL-SMOTE and BL-CLBN maintain the close proximity to the real data PDF. On the other hand, the distance of BL-GAN from the real distribution becomes even larger.

2) Pairwise Correlation Comparison Results: To avoid cluttering, Fig. 7 shows only the best and worst correlation which are from SMOTE and BL-GAN, respectively. Results from Fig. 7 show that the correlation between variables in real and synthetic data set is maintained using SMOTE. This is indicated by smaller values of pairwise correlation for SMOTE. However, pairwise correlation difference is high for BL-GAN, i.e, the relationship between source RSRP and source G-Factor is 0.67 as well as the correlation between target RSRP and target G-Factor is 0.36. These results show that BL-GAN does not yield synthetic data in which the relationship between the input features mimics that in real data.

3) Synthetic Data Evaluation Summary: Summary of results evaluation in Table IV show that SMOTE and CLBN, together with their BL counterpart, generate data with good quality in terms of approximating the distribution and feature correlation with respect to the real data. The corresponding small values of ζ for these techniques signify that the marginal distributions of real and synthetic data sets are very close to each other. Similarly, the small values of μ indicate the unchanged relationship between the input features. On the other hand, the quality of data generated by GAN and BL-GAN is poorer compared to other techniques. The poor performance of GAN can be attributed to the fact that it is based on deep learning, which requires a lot of training data unlike SMOTE and CLBN. This evaluation shows that GAN is not suitable for addressing imbalance data problem with few training samples.

Ideally, data with poor quality should be discarded and should not be utilized for AI model training. Such kind of data might mess up with model training instead of improving them. However, to show the impact on model training, we keep the data generated by GAN to observe how it affects the model performance.

B. Handover Failure Prediction Results

To evaluate the effectiveness of the synthetic data generation in improving the performance of our HOF prediction model, we analyze the AUROC and AUPRC of different AI algorithms with and without data augmentation. Fig. 8(a) shows the AUROC comparison of different AI models. We observe that data augmentation techniques such as SMOTE and BL-SMOTE do not provide a significant improvement in terms of AUROC. AUROC of the model trained on the data without augmentation has an average of 0.936, which is similar to the average AUROC of models trained with augmented data from SMOTE and BL-SMOTE. Meanwhile, training the models using data generated from CLBN and BL-CLBN show improvement with an average AUROC of around 0.941 for both. When models are trained using data from GAN, average AUROC is again comparable when using real data only. However, using BL-GAN, AUROC drops to 0.928. Results also show that augmentation technique is particularly beneficial for Random Forest and KNN.

Results in Fig. 8(b) shows the variations in the AUPRC for different models trained with imbalance data set and balance data set using different augmentation techniques. Evaluation of the models which are trained using SMOTE and BL-SMOTE indicates that these approaches cause degradation in the performance of the classifiers. From AUPRC of 0.652 without data augmentation, AUPRC drops to 0.629 and 0.643 with application of SMOTE and BL-SMOTE respectively. Particularly, SMOTE and BL-SMOTE cause a significant decline in the performance of Random Forest and XGBoost, which drags down the total average AUPRC. Meanwhile, models which utilize data generated from CLBN and BL-CLBN for training show AUPRC improvement. Both methods seem to work well in improving AUPRC of almost all the AI classifiers especially KNN and Random Forest. Improvement in the performance of AdaBoost is also notable using CLBN-based augmentation techniques. However, once again, GAN and BL-GAN fail to enhance the overall AUPRC performance of the models. These techniques only improve the performance of KNN while the rest of the classifiers degrade. These results mirror the poor quality of data generated using GAN.

These results highlight the importance of thorough synthetic data evaluation. Without data evaluation, models train with GAN-based data augmentation seem to perform fine. In fact, utilizing GAN even improved the overall AUROC of the models. These results reveal that looking solely on the effect of data augmentation techniques on prediction performance of the model is tricky and extra precaution is needed when using data augmentation techniques such as GAN. Results also show that different AI models have different sensitivities to the quality of data fed into them. For instance, the performances of Support Vector Classifier and AdaBoost are severely affected by the poor quality of data while others are not.

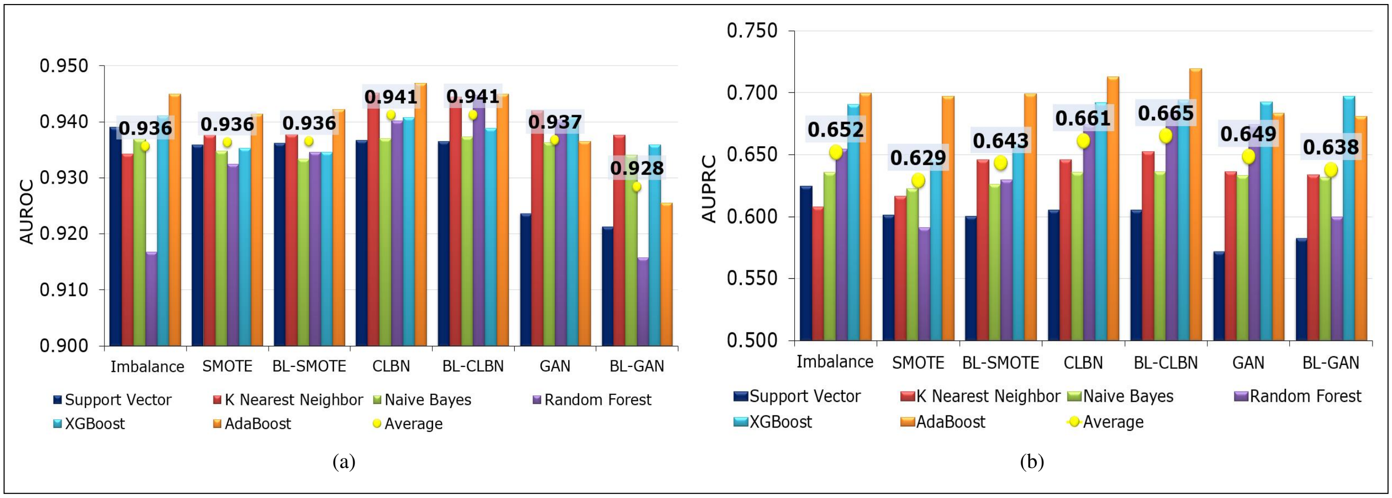


Fig. 8. AUROC (a) and AUPRC (b) comparison between different augmentation techniques using different AI algorithms.

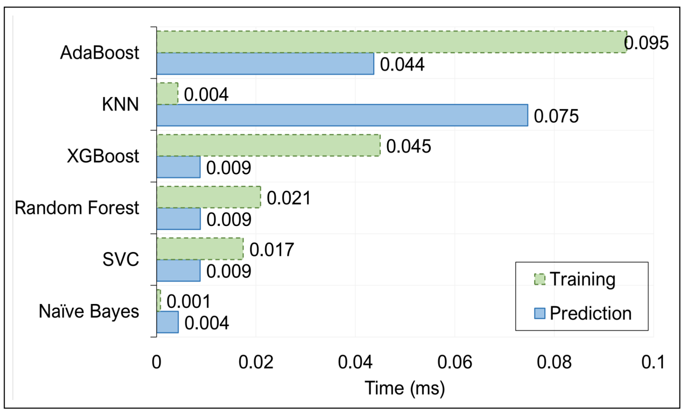


Fig. 9. Time complexity comparison of different AI models.

We also evaluate the time complexity of the AI models we used for HOF prediction. Time complexity is an important performance metric especially for time sensitive problems such as HOF prediction. Fig. 9 shows the training and prediction time of the AI models normalized against the number of data points in training and test set. In terms of training time, results show that Naïve Bayes and KNN are the fastest while AdaBoost is the slowest. Meanwhile, with regards to the prediction time, again, Naïve Bayes has the shortest time while KNN has the longest followed by AdaBoost. However, although faster, the prediction accuracy of algorithms such as Naïve Bayes is significantly worse than AdaBoost. Moreover, the prediction time of AdaBoost, although higher than the others, is significantly less than the least possible TTI of 5G communication. 5G design includes scalable TTI, corresponding to slot duration between 62.5μs and 1ms. In 3GPP Release 16, 120kHz is the maximum sub-carrier spacing allowed for data communication which makes the lowest achievable slot duration equal to 125μs, far longer than the 44μs prediction time of AdaBoost. Thus, even with high complexity, we use AdaBoost to leverage the superior performance in detecting HOFs.

C. Shift in the Decision Boundary after Class Balancing

The shift in the decision boundary towards the minority class after data augmentation poses a trade-off between detection of more HOFs and the overall accuracy of the model. This trade-off reflects with the variation in the confusion matrix using different data augmentation techniques as shown in Fig. 10. The confusion matrix results are shown only for the best performing classifier, AdaBoost, using the validation data set. Results reveal that AdaBoost model correctly classify 20 out of 36 HOFs when it is trained on the imbalance data set with only 7 HOS misclassified as HOF. Meanwhile, the number of correctly detected HOFs increase to 31 with the utilization of BL-SMOTE. However, the number of misclassified HOS as HOF increase from 7 to 64 alongside. Similarly, the HOF detection performance of the model trained with synthetic data from BL-GAN increases in expense of more HOS misclassifications. On the contrary, BL-CLBN displays a fair detection rate of HOF without causing too many misclassifications of HOS as HOF. BL-CLBN correctly predict 28 out of 36 HOFs with 22 HOS misclassified as HOF.

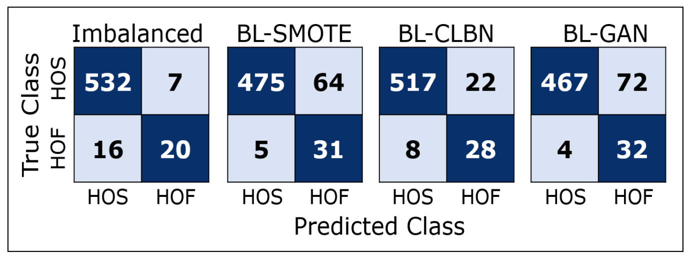


Fig. 10. Confusion matrix comparison between different data augmentation techniques using AdaBoost

D. Handover Failure Mitigation Results

The effectiveness of the proposed HOF mitigating scheme hinges on the performance of the HOF prediction model. Having analyzed the performances of various HOF prediction models using different class balancing techniques, in this subsection we discuss the results of handover failure mitigation algorithm.

Fig. 11 shows the performance of the HOF mitigation algorithm using the percentage of HOF resolved with corresponding false compensation rate. Results show that without

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3157802, IEEE Transactions on Vehicular Technology

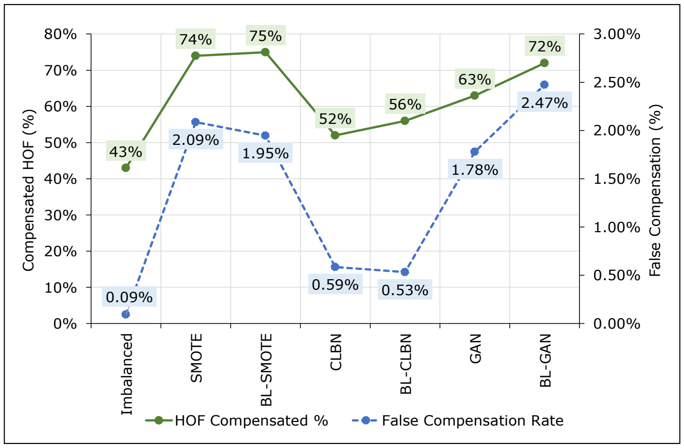


Fig. 11. Handover failure mitigation performance using different data augmentation techniques.

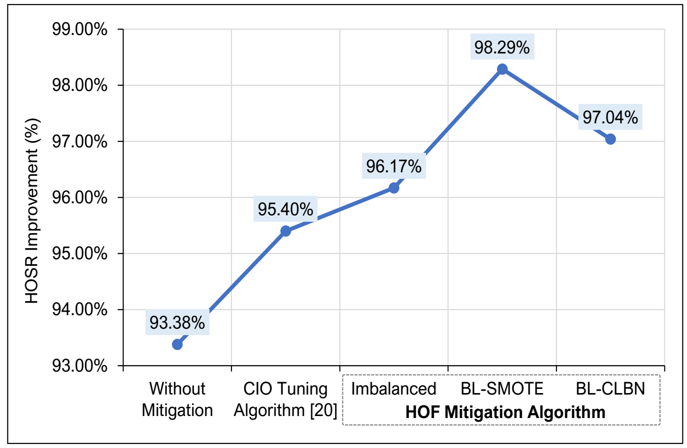


Fig. 12. HOSR ξ improvement comparison between algorithm used in [20] and proposed HOF mitigation algorithm.

TABLE V
FAILED COMPENSATION AND FALSE MITIGATION SCORE Υ FOR DIFFERENT VALUES OF α

	$\alpha = 0.5$	$\alpha = 0.1$	$\alpha = 0.9$
Imbalanced	0.500	0.900	**0.100**
SMOTE	0.435	0.113	0.757
BL-SMOTE	0.389	**0.078**	0.701
CLBN	0.464	0.670	0.259
BL-CLBN	**0.388**	0.550	0.225
GAN	0.535	0.395	0.674
BL-GAN	0.541	0.174	0.908

any data augmentation, around 43% of the HOFs are mitigated. However, this number increases when data augmentation is incorporated. For instance, the highest number of HOFs are detected using SMOTE and BL-SMOTE. Thus, more HOFs are compensated by the HOF mitigation algorithm which is around 30% more compared to imbalanced data set. However, it is notable that using SMOTE and BL-SMOTE will also lead to a higher false compensation rate of around 2.09% and 1.95%, respectively compared to 0.09% for imbalanced data set. A more conservative choice is either CLBN or BL-CLBN with only 0.59% and 0.53% false compensation rate, respectively. At around 56%, the performance of BL-CLBN in terms of resolving HOFs is better compared to CLBN at 52%.

Table V shows the failed compensation and false mitigation score (Υ) given in eq. (10) with different values of α. We observe that a model trained on BL-CLBN generated synthetic data set performs the best when equal priority (α = 0.5) is given for uncompensated HOF and false compensation. Meanwhile, model utilizing data using BL-SMOTE has the best performance when more weight is given to minimize false compensation rather than failed compensation (α = 0.1). Model trained without data augmentation performs the best when more weight is given to minimize failed HOF compensation (α = 0.9). These results show the capability of the proposed HOF mitigation approach to provide leeway to the network operators to set priorities as well as select the appropriate data augmentation technique.

Thus far, it is evident that the current solution offers two mechanisms to benefit the insights gained from the false compensation analysis: (1) A tunable weight α is incorporated wherein a vendor or operator can choose to minimize the false compensation at the expense of missing more handover failure detection as discussed in Section VI, and (2) The results in Fig. 8, Fig. 11, and Table V provide insights for operators to choose an AI and augmentation method combo that offers that aligns with operators' priority for minimizing HOF failures or minimizing unwanted momentary increase transmission power. These methods offer an offline strategy to minimize the false compensation at the expense of less HOF mitigation.

Finally, to verify the effectiveness of the proposed TORIS framework, we compare its performance with the state-of-the-art alternative, i.e., CIO tuning algorithm [20], for improving HOSR. The algorithm used by [20] is implemented in SyntheticNET [31]. For a fair comparison, the same network deployment, number of users, user speed, user mobility as well as the same handover parameters are used. Fig. 12 presents the findings of the comparison study. The results show that while the CIO-based tuning to mitigate HOF improves the HOSR from 93.38% to 95.40%, the proposed TORIS improves the HOSR to around 96%-98%. It is worth noting that comparison in terms of false mitigation could not be done as proposed solution in [20] does not involve variation in Tx power.

VIII. CONCLUSION AND FUTURE WORK

We propose an AI powered Tx power tuning-based solution to improve inter-frequency handover success rate, called TORIS. TORIS consists of two distinct components. 1) An AI-based model to accurately predict HOF despite of imbalanced training data. 2) A serving and target cell Tx power tuning algorithm that feeds on the prediction of the HOF model to minimize the HOF and thus improve HOSR. First challenge in creating an HOF model is selection of training feature. We leverage domain knowledge to devise novel set of training features that outperform the feature combinations used in literature. The second key challenges in creating an AI-based HOF prediction model comes from the imbalanced training data where the number of HOF instances in real network are far less than the HOS. We show that this data imbalance can lead to poorly performing model even with the state-of-the-art classifiers. We solve this problem by thoroughly investigating a range of data augmentation techniques and their novel combinations such as SMOTE, CLBN and GAN. Leveraging the idea of using the borderline, we devise a hybrid CLBN and GAN called BL-CLBN and BL-GAN. Results show that the former performs best in terms of AUROC and AUPRC of the HOFs model trained on the augmented data. Extensive synthetic data evaluation shows that the quality and validity of data generated by GAN is poorer in comparison with data generated using other techniques due to its inherent dependency on large training data. Both SMOTE based and GAN based approaches tend to shift the models' decision boundaries toward detection of more HOFs at expense of misclassification of HOSs as HOFs. On the other hand, CLBN based HOF prediction models predict more HOFs compared to imbalanced data trained HOF models while keeping in check the amount of misclassified HOSs. Upon quantifying the effect of misclassifications, BL-CLBN shows superior performance in mitigating HOFs when equal priority is given to HOF prediction and HOS misclassifications. Finally, we compare the performance of TORIS with state-of-the-art HOSR enhancement scheme that uses CIO tuning instead of Tx power tuning. Results show that proposed TORIS scheme outperforms the state-of-the-art by significant margin in improving the HOSR i.e., CIO based scheme raises HOSR from 93.38% to only 95.40%. In contrast TORIS can raise HOSR all the way to 98.29%. The gain of the proposed TORIS can be attributed to novel feature selection, data balancing and parameter turning algorithm that addresses the root cause of the HOF i.e., Tx power misconfiguration instead of tuning CIO.

For future work, we will investigate other types of AI models such as anomaly detectors for HOF prediction. Additionally, we will devise a mechanism to counter the false compensation by using a reinforcement learning based HOF mitigation approach. Using this approach, we can analyze the impact of reducing Tx power to the overall performance of the network. To make the proposed solution more holistic, for future work, we also aim to perform a joint improvement of intra-frequency and inter-frequency HOs.

Acknowledgment

This work is supported by the National Science Foundation under Grant Numbers 1718956 and Qatar National Research Fund (QNRF) under Grant No. NPRP12-S 0311-190302. The statements made herein are solely the responsibility of the authors. For more details about these projects please visit: http://www.ai4networks.com

References

[1] A. Imran, A. Zoha, and A. Abu-Dayya, "Challenges in 5G: how to empower SON with big data for enabling 5G," *IEEE Network*, vol. 28, pp. 27–33, Nov 2014.

[2] Syed Muhammad Asad Zaidi, Marvin Manalastas, Hasan Farooq, and A. Imran, "Mobility Management in 5G and Beyond: A Survey and Outlook," *IEEE Access*, 2020.

[3] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, "Ultra-reliable and low-latency communications in 5G downlink: Physical layer aspects," *IEEE Wireless Communications*, vol. 25, no. 3, pp. 124–130, 2018.

[4] K. Dimou, M. Wang, Y. Yang, M. Kazmi, A. Larmo, J. Pettersson, W. Muller, and Y. Timner, "Handover within 3GPP LTE: Design principles and performance," in *2009 IEEE 70th Vehicular Technology Conference Fall*, pp. 1–5, 2009.

[5] "Fujitsu Whitepape: Enhancing LTE Cell-Edge Performance via PDCCH ICIC."

[6] M. U. B. Farooq, M. Manalastas, W. Raza, A. Ijaz, S. M. A. Zaidi, A. Abu-Dayya, and A. Imran, "Data driven optimization of inter-frequency mobility parameters for emerging multi-band networks," 2020.

[7] M. U. B. Farooq, "Data driven optimization of inter-frequency mobility parameters for emerging networks," *Master's Thesis, University of Oklahoma-Tulsa*, 2021.

[8] M. U. B. Farooq, M. Manalastas, S. M. A. Zaidi, A. Abu-Dayya, and A. Imran, "Machine learning aided holistic handover optimization for emerging networks," in *2022 IEEE International Conference on Communications (ICC)*, IEEE, 2022.

[9] M. Manalastas, "Realistic modeling of handover events in a multi-carrier 5G network: A preliminary step towards COP-KPI relationship realization," *Master's Thesis, University of Oklahoma-Tulsa*, 2020.

[10] E. U. T. R. Access, "Self-configuring and self-optimizing network use cases and solutions," *Protocol specification (Release 9)*, 2011.

[11] S. Khunteta and A. K. R. Chavva, "Deep learning based link failure mitigation," in *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 806–811, 2017.

[12] Xiaomeng Shu, Li Zhu, Hongli Zhao, and Tao Tang, "A novel hand-off decision algorithm in TD-LTE based train-ground communication system," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 757–761, 2014.

[13] H. Qu, Y. Zhang, J. Zhao, G. Ren, and W. Wang, "A hybrid handover forecasting mechanism based on fuzzy forecasting model in cellular networks," *China Communications*, vol. 15, no. 6, pp. 84–97, 2018.

[14] A. Kulkarni, A. Seetharam, A. Ramesh, and J. D. Herath, "Deepchannel: Wireless channel quality prediction using deep learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 443–456, 2020.

[15] P. Muñoz, R. Barco, and I. de la Bandera, "On the potential of handover parameter optimization for self-organizing networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 1895–1905, 2013.

[16] D. Castro-Hernandez and R. Paranjape, "Optimization of handover parameters for LTE/LTE-A in-building systems," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 6, pp. 5260–5273, 2017.

[17] X. Cai, C. Wu, J. Sheng, J. Zhang, and Y. Wang, "A parameter optimization method for LTE-R handover based on reinforcement learning," in *2020 International Wireless Communications and Mobile Computing (IWCMC)*, pp. 1216–1221, 2020.

[18] A. S. Priyadharshini and P. T. V. Bhuvaneswari, "A study on handover parameter optimization in LTE-A networks," in *2016 International Conference on Microelectronics, Computing and Communications (MicroCom)*, pp. 1–5, 2016.

[19] K. Kitagawa, T. Komine, T. Yamamoto, and S. Konishi, "A handover optimization algorithm with mobility robustness for LTE systems," in *2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1647–1651, 2011.

[20] J. Joseph, F. Ahmed, T. Jokela, O. Tirkkonen, J. Poutanen, and J. Niemela, "Big data enabled mobility robustness optimization for commercial LTE networks," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, IEEE, 2020.

[21] Z. Ali, N. Baldo, J. Mangues-Bafalluy, and L. Giupponi, "Machine learning based handover management for improved QoE in LTE," in *NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium*, pp. 794–798, 2016.

[22] J. Xu, Y. Zhao, and X. Zhu, "Mobility model based handover algorithm in LTE-Advanced," in *2014 10th International Conference on Natural Computation (ICNC)*, pp. 230–234, 2014.

[23] K. S. Lee and A. Park, "Reduction of handover failure for small cells in heterogeneous networks," in *2014 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 707–708, 2014.

[24] M. B. Gharbia and R. Bouallegue, "Handover decision algorithm in femtocell long term evolution networks," in *2018 Seventh International Conference on Communications and Networking (ComNet)*, pp. 1–6, 2018.

[25] Y. Mal, J. Chen, and H. Lin, "Mobility robustness optimization based on radio link failure prediction," in *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*, pp. 454–457, 2018.

[26] 3GPP, "Technical Specification Group Radio Access Network; NR; NR and NG-RAN Overall Description; Stage 2," TS 38.300, 3rd Generation Partnership Project (3GPP). V16.5.0 (Release 16).

[27] P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, "A survey of machine learning techniques applied to self-organizing cellular networks,"

IEEE Communications Surveys & Tutorials, vol. 19, no. 4, pp. 2392–2431, 2017.

[28] H. Wang, Z. Wu, S. Ma, S. Lu, H. Zhang, G. Ding, and S. Li, "Deep learning for signal demodulation in physical layer wireless communications: Prototype platform, open dataset, and analytics," *IEEE Access*, vol. 7, pp. 30792–30801, 2019.

[29] M. S. Mollel, A. I. Abubakar, M. Ozturk, S. F. Kaijage, M. Kisangiri, S. Hussain, M. A. Imran, and Q. H. Abbasi, "A survey of machine learning applications to handover management in 5g and beyond," *IEEE Access*, vol. 9, pp. 45770–45802, 2021.

[30] M. Tayyab, G. P. Koudouridis, and X. Gelabert, "A simulation study on LTE handover and the impact of cell size," in *Broadband Communications, Networks, and Systems* (V. Sucasas, G. Mantas, and S. Althunibat, eds.), (Cham), pp. 398–408, Springer International Publikoularidou,shing, 2019.

[31] S. M. A. Zaidi, M. Manalastas, H. Farooq, and A. Imran, "Syntheticnet: A 3GPP compliant simulator for AI enabled 5G and beyond," *IEEE Access*, pp. 1–1, 2020.

[32] K. W. Bowyer, N. V. Chawla, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *CoRR*, vol. abs/1106.1813, 2011.

[33] J. Young, P. Graham, and R. Penny, "Using bayesian networks to create synthetic data," *Journal of Official Statistics*, vol. 25, no. 4, p. 549, 2009.

[34] C. Chow and C. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 462–467, 1968.

[35] N. Patki, R. Wedge, and K. Veeramachaneni, "The synthetic data vault," in *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 399–410, Oct 2016.

[36] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, "Modeling tabular data using conditional GAN," in *Advances in Neural Information Processing Systems*, 2019.

[37] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-smote: A new over-sampling method in imbalanced data sets learning," in *Advances in Intelligent Computing* (D.-S. Huang, X.-P. Zhang, and G.-B. Huang, eds.), (Berlin, Heidelberg), pp. 878–887, Springer Berlin Heidelberg, 2005.

[38] G. E. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD explorations newsletter*, vol. 6, no. 1, pp. 20–29, 2004.

[39] J. Zhang and X. Li, "Phishing detection method based on borderline-smote deep belief network," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*, pp. 45–53, Springer, 2017.

[40] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017.

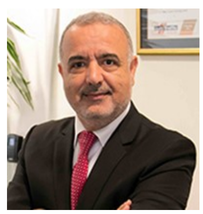
[41] A. Goncalves, P. Ray, B. Soper, J. Stevens, L. Coyle, and A. P. Sales, "Generation and evaluation of synthetic patient data," *BMC Medical Research Methodology*, vol. 20, pp. 1–40, 2020.

[42] I. Sobol, "Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates," *Mathematics and Computers in Simulation*, vol. 55, no. 1, pp. 271 – 280, 2001. The Second IMACS Seminar on Monte Carlo Methods.

Marvin Manalastas received a B.S. degree in electronics and communication engineering from the Polytechnic University of the Philippines in 2011, and an M.S. degree in electrical and computer engineering from The University of Oklahoma in 2020. He is currently pursuing a Ph.D. degree in electrical engineering with The University of Oklahoma, working AI4Networks Research Center. He has more than seven years of industry experience in the field of cellular network optimization in the Philippines and Japan. He did multiple internships in the USA and worked as an RF Optimization intern in Mobilecomm Professionals, TX, as an AI/ML intern for Synopsys, VA, and as a research fellow at US FDA, MA. His research interests include machine learning applied in the optimization of 5G and beyond networks.

Muhammad Umar Bin Farooq received the BS degree in electrical engineering from Lahore University of Management Sciences in 2015 and MS degree in computer science from Information Technology University, Pakistan in 2018. He received another MS degree in electrical and computer engineering from University of Oklahoma in 2021. He is currently pursuing a Ph.D. degree in electrical and computer engineering at University of Oklahoma, where he is a part of AI4Networks research center. He has done multiple internships in AT&T Labs and Interdigital. His research interests include machine learning based zero-touch optimization for 5G and beyond networks.

Syed Muhammad Asad Zaidi received the B.Sc. degree in information and communication engineering from the National University of Science and Technology (NUST), Pakistan, in 2008, MS from Ajou University, South Korea in 2013, and PhD Electrical engineering from University of Oklahoma in 2021. With almost 13 years' experience in telecom industry, he has worked in Mobilink, Pakistan, Korea Electronics and Technology Institute (KETI), South Korea, MOTiV Research, Japan, AT&T, USA, Sprint, USA and T-Mobile, USA. Currently, he is leading 5G radio-frequency optimization team in T-Mobile networks. His research domain is mobility robustness and optimization of futuristic ultra-dense base station deployment.



Adnan Abu-Dayya (Senior Member, IEEE) received the Ph.D. degree in digital mobile communications (electrical engineering) from Queen's University, Kingston, ON, Canada, in 1992. He worked with ATT Wireless, Seattle, USA, for ten years, where he served in a number of senior management positions covering product innovations, emerging technologies, systems engineering, product realization, and intellectual property management. He is the Executive Director of Qatar Mobility Innovations Center, Doha, Qatar. He led the establishment of the Qatar Mobility Innovations Center in 2009. It is one of the first independent innovations institution in the Middle East focused on translating Research and Development and technology innovations into scalable digital platforms and solutions in the field of intelligent mobility and smart cities. He worked as a Senior Manager with the Advanced Technology Group, Nortel Networks, Canada, and as a Senior Consultant with the Communications Research Centre, Ottawa, ON, Canada. He has ten issued patents, and about hundred refereed publications. Dr. Abu-Dayya serves as the Chairman of the Advisory Board of the Electrical and Computer Engineering Department, Texas AM University at Qatar, where he is a member of the Steering Committee of the Smart Grid Research Center.

Ali Imran (Senior Member, IEEE) is a Presidential Associate Professor and Willams Professor in ECE and founding director of the Artificial Intelligence (AI) for Networks (AI4Networks) Research Center and TurboRAN Testbed for 5G and Beyond at the University of Oklahoma. His research interests include AI and its applications in wireless networks and healthcare. His work on these topics has resulted in several patents and over 100 peer-reviewed articles including some of the highly influential papers in the domain of wireless network automation. On these topics he has led numerous multinational projects, given invited talks/keynotes and tutorials at international forums and advised major public and private stakeholders and co-founded multiple start-ups. He holds a B.Sc. degree in electrical engineering from the University of Engineering and Technology Lahore, Pakistan, in 2005, and the M.Sc. degree (Hons.) in mobile and satellite communications and the Ph.D. degree from the University of Surrey, Guildford, U.K., in 2007 and 2011, respectively. He is an Associate Fellow of the Higher Education Academy, U.K. He is also a member of the Advisory Board to the Special Technical Community on Big Data, the IEEE Computer Society.