Failure Prediction for Proactive Beam Recovery in Millimeter-Wave Communication

Ayah Abusara[1], Hossam S. Hassanein[2], Aboelmagd Noureldin[3] and Akram Bin Sediq[4]

[1]Electrical and Computer Eng. Dept., Queen's University, Kingston, Canada, 18ama23@queensu.ca
[2]School of Computing, Queen's University, Kingston, Canada, hossam@cs.queensu.ca
[3]Electrical and Computer Eng. Dept., Royal Military College of Canada, Kingston, Canada, noureldin-a@rmc.ca
[4]Ericsson, Ottawa, Canada, akram.bin.sediq@ericsson.com

ICC 2022 - IEEE International Conference on Communications | 978-1-5386-8347-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICC45855.2022.9838772

***Abstract*—This paper proposes beam failure prediction to recover from inevitable link failures in beam-based mmWave communication proactively. The proposed system consists of two components. First, a prediction engine to foresee future beam failures and their severity. For this purpose, machine learning and deep learning are proposed to perform prediction. The second component is a proactive recovery mechanism, that matches the prediction failure results with a suitable recovery action, with the goal to maintain seamless connectivity and prevent service interruptions. The performance of the proposed system is compared against conventional beam failure detection and recovery. Simulations were carried out using real beamforming data. The results indicate a substantial improvement in the network performance. The improvement is measured in terms of prediction accuracy, beam failure probability and successful beam failure probability. This paper also assesses a drawback of the proposed system, particularly the increase in handover rate, and shows that the achieved gain outweighs this weakness.**

***Index Terms*—Millimeter Wave, Beamforming, Beam Management, Beam Failure Prediction, Beam Failure Recovery, Machine Learning, Deep Learning**

I. Introduction

Exploiting the millimeter wave (mmWave) band has been a game-changer for the telecommunications industry. Ideal for achieving unprecedented data rates and record-breaking capacity, mmWave is a cornerstone technology driving the revolution in cellular networks. The potential of mmWave lies in operating at Extremely High Frequencies (EHF), i.e., *30-300 GHz*, which offers large amounts of untapped bandwidth to users. However, communication in the mmWave band is subject to significant path-loss, atmospheric and precipitation attenuation, blockages, scattering and diffraction [1]–[3]. Due to these constraints, the range of mmWave is limited [4], [5].

Fortunately, massive multiple-input multiple-output (Massive MIMO) and beamforming technology can address those challenges by using many antenna arrays and highly directional beams. A large number of antennas is needed to achieve the required array gain, while the directional beams are to focus the transmitted energy so that the signals are steered away from obstacles, thus extending the communication range of mmWave [2], [3], [6]. Despite its advantages in realizing mmWave, beam-based communication imposes several new challenges on the network. In the presence of the dynamically changing network conditions and user mobility, the network has to maintain perfect beam alignment between the user equipment (UE) and the next generation node base station (gNB) to maintain reliable connectivity [2], [7]. Several beam management (BM) functions are adopted to facilitate beam-based communication. Amongst these are beam failure detection and recovery functions, which are vital to support seamless connectivity on the beamformed mmWave band. Beam failure detection and recovery are needed to overcome deafness and blockages resulting from user mobility and less than ideal propagation conditions [8]. The primary function of beam failure detection is to constantly monitor the radio link quality (RLQ) of serving beams in search of an indication of failure. When failure is detected, beam failure recovery is activated, triggering the process to regain fast access to another serving link [7].

Despite the extensive research work and standardization effort in this area, many questions remain regarding the detailed workflow of beam management in general and beam failure detection and recovery in specific [7]. Moreover, the research and the standardization on failure detection and recovery have focused mainly on static users and environments, whereas there have been few discussions addressing the non-ideal conditions [9]. Also, research to date has not been able to confirm how failure detection and recovery will be integrated with existing radio resource management (RRM) functions [7]. Some studies have contributed to this area, the study in [5] explores the issue of beam misalignment for self-healing unmanned aerial vehicle (UAV) networks. The authors propose a medium access control (MAC) mechanism to recover from failures by finding an alternative link that bypasses the faulty UAV interrupting the communication in a mmWave fly-mesh [5]. In [1], the authors propose to use diversity coding for multi-beam transmission to avoid service interruptions due to blockages. Alternatively, machine learning (ML) and intelligent reflecting surfaces (IRS) are proposed in [9] to overcome the challenging propagation conditions inherent in the mmWave band, thereby reducing failure incidents. In [10], the authors propose using the angular domain information (ADI) to enhance the agility of BM functions.

This paper makes an original contribution to improve beam failure detection and recovery for next generation cellular networks (NGCN). The paper focus on mmWave deployment

Authorized licensed use limited to: Queen's University. Downloaded on September 18,2022 at 18:22:22 UTC from IEEE Xplore. Restrictions apply.

in challenging dense urban environments. Moreover, this paper considers a specific beam failure case, that is when all beams fail at the same moment. Motivated by the self organizing networks (SON) concept, specifically self-healing networks, this paper proposes automating the failure detection and recovery to ensure seamless connectivity on the mmWave band in NGCN. For this purpose, deep learning (DL) is used to forecast beamforming failure events and prepare for recovery prior to the failure.

The rest of the paper is organized as follows: Section II describes the problem and defines the proposed system. Section III provides an overview and analysis of the dataset on which this paper is based. Then in Section IV-B, the simulations and results are discussed. Conclusions and future work are discussed in Section V.

II. Problem Formulation and System Model

In NGCN, the radio access network (RAN) will be composed of many small cells (i.e., micro-cells, pico-cells, femto-cells and D2D), overlaying the legacy macro cells. Both mmWave and dense small cells are used to enable the speed and capacity goals of NGCN. In order to operate efficiently, BM functions are needed in small cell deployment where beamformed mmWave is put in use. BM functions include beam sweeping, beam tracking, beam switching, beam failure detection and recovery, see Fig. 1. As illustrated in the figure, beam sweeping is used for initial access, and it refers to a full-space beam search [7]. Beam tracking is used to ensure that the UE is connected to the beam with the best signal throughout the connection session. The beam quality is assessed using the reference signal received power (RSRP) measured and reported by the UE to the base station. Beam switching takes place when the reported RSRP of the serving beam drops below the RSRP of another beam. If the link quality drops below a certain threshold, depending on predefined failure criteria, failure is declared, and recovery is triggered [7]. In some instances, beam recovery will fail to reassign another link to the UE, if there is a delay in getting a response back from the candidate links [7]. Another cause for a recovery failure is when all candidate beams are equivalently failing, i.e., when the UE detects no signal from any of the beams which is the interest of this paper. If recovery fails, the connection will be interrupted, and a radio link failure (RLF) will be declared. Therefore, passive beam failure detection and recovery are insufficient to ensure agile and reliable communication over the mmWave band.

A. Proposed Model

In this section, the proposed self-healing beamformed mmWave system is presented. The beam failure prediction and proactive recovery, achieving self-healing, are depicted in Fig. 2. The proposed system is composed of two parts. The first part is responsible for beam failure prediction. For prediction, offline training is needed for the prediction model to learn dependencies and underlying patterns in the data. For this purpose, a historical beamforming dataset with location and

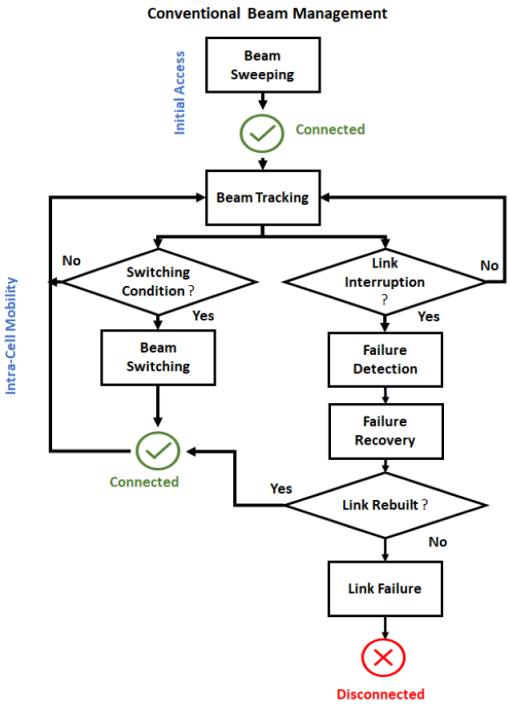

Fig. 1. Conventional passive beam management procedure

time fingerprints is used. Different ML and DL algorithms are examined for this problem. The algorithms used in this paper are defined briefly below:

- Multi-layer perceptron (MLP): is a class of feed-forward artificial neural networks. MLP is a powerful tool that attempts to mimic the human brain in learning complex input/output relationships. Thus, it is a supervised learning algorithm. It breaks complex models into layers of small functions. In this paper, MLP is used as a baseline model. Two hidden layers with 20 nodes each are used to construct the MLP learner.
- XGBoost: is a well-known ensemble model that combines different models to classify a set of records. The different models are expected to make uncorrelated errors, which can be averaged out to achieve better total accuracy. In this paper, XGBoost is also used as a baseline model. To build XGBoost, a total of 20 models are used, and decision trees (DTs) are chosen to be the booster.
- Long short-term memory (LSTM): is a sequential prediction network. Those networks have loops and memory cells that enable them to memorize past values. In particular, LSTM can memorize long-term dependencies in the data sequence. In this paper, one input layer, one LSTM layer with 20 elements and one dense output layer are used to build the LSTM learner.

The trained prediction model then can be used to predict future beam failure. At each time step, the UE location and the previous beam strength measurements are used as inputs to the prediction model. The predicted output conveys two things: failure flag (0 or 1) and anticipated failure duration, measured

Authorized licensed use limited to: Queen's University. Downloaded on September 18,2022 at 18:22:22 UTC from IEEE Xplore. Restrictions apply.

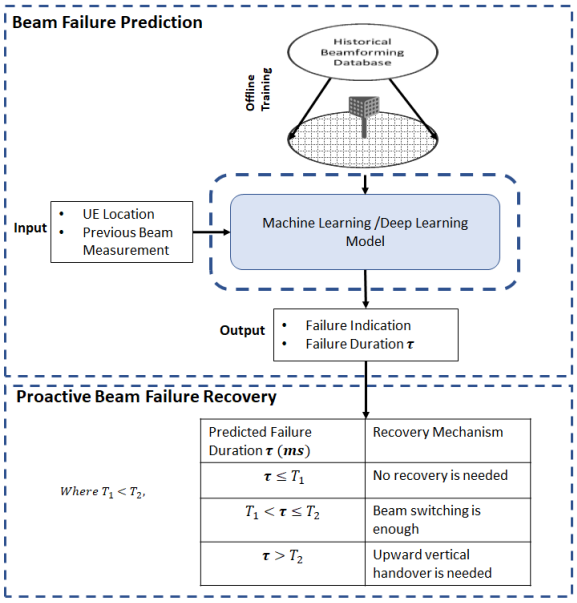

Fig. 2. System model for beam failure prediction and proactive beam recovery

in milliseconds (*ms*). The latter output is important since it implies the first output, and by the duration, it predicts the severity of the anticipated failure.

This takes the discussion to the second component of the proposed system, which defines three possible recovery mechanisms based on the predicted failure duration. The defined mechanisms are no action, beam switching and vertical handover (HO), here vertical HO refers to a user transitioning between cells belonging to different network technologies [11]. Based on two pre-configured time thresholds T_1 and T_2, where $T_1 < T_2$. T_1 defines the maximum allowable time a signal can be absent with negligible effect on the communication quality (all beams measure -1). In this paper, T_1 is assumed to be *200 ms* and T_2 is set to *600 ms*. This assumption of the T_1 value is fair, given transmitting beam bursts for beam acquisition requires between *5 to 160 ms* [3]. Hence, *200 ms* of signal absence will be tolerated. Also, a signal absence for more than *600 ms* is substantial and might affect the connection quality, hence, it is considered a severe failure requiring handover to the macro cell (i.e., vertical handover). Any failure duration between the two thresholds is considered mild, requiring beam switching.

III. Data Acquisition and Analysis

A. Beamforming Dataset

The source of the dataset used in this paper is an experimental gNB located at Ericsson, Lund, Sweden. The UE is served by wide beams used mainly for transmitting synchronization signal blocks (SSB), and narrow beams used for data transmission. Ericsson conducted a series of measurement experiments to assess the wide and narrow beam RSRP reported to the gNB by a UE. Data collection was performed multiple times on different days while going over the route shown in Fig. 3. There are two notable features about this route, it is located in a dense urban area, and it is mostly non-line of sight (NLoS). To avoid sample bias, beam strength data were collected under different speed conditions; slow walking, fast walking and recreational bicycling. In addition, different movement directions, UE orientations and UE mounting positions were used in order to increase randomness in the data. The dataset consists of RSRP time and location fingerprints. The RSRP reported by the UE was recorded for every timestamp, in parallel with recording the GPS location. The reported RSRP is for different beams, narrow beams and wide beams. In line with 3GPP standards, an indication of RSRP measurement is used instead of the negative dBm RSRP values. Hence, the reported values are integers ranging from *16 to 113*; and the mapping between these indications and the corresponding RSRP range is given in [12, Tab.10.1.6.1-1]. The value of *-1* is used to indicate no signal. In this paper, a full-beam failure event is studied. More specifically, a full-beam failure event is labeled when all beam links measure *-1* by the UE.

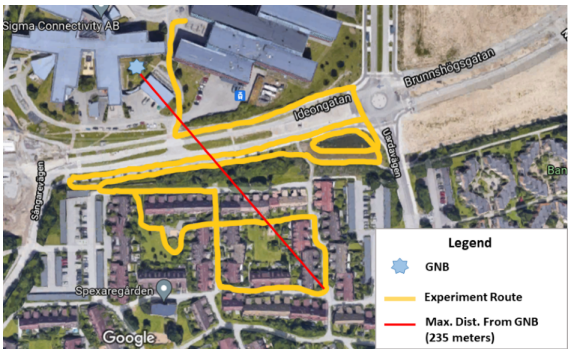

Fig. 3. Map of the experimental setup

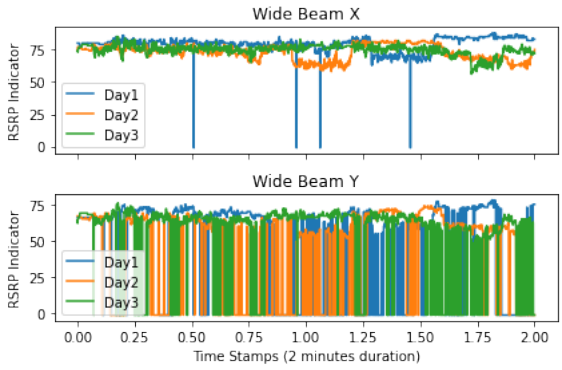

Fig. 4. Temporal analysis of the signal strength reported from two different wide beams *X & Y*, for three different days at the same location

B. Data Analysis

With the dataset in hand, the temporal and spatial properties are obtained.

1) Temporal Analysis: Fig. 4 shows the strength measurements of two distinct wide beams: *X* and *Y*. The strength values were collected from the same location point for a 2

Authorized licensed use limited to: Queen's University. Downloaded on September 18,2022 at 18:22:22 UTC from IEEE Xplore. Restrictions apply.

minute time window on three different days. It is clear from the graph that the strength measurements exhibit strong temporal patterns. To illustrate, the reported RSRP is about 80, which indicates a good signal received on beam *X*. This signal was almost consistent on the other days except for a slight variation due to dynamic unregulated variables during the experiment (e.g., the weather and the dynamic environment). Temporal information also indicated if a beam signal is unstable, as was the case of beam *Y*. Moreover, autocorrelation is calculated in order to estimate the correlation between the beam strength values with values from previous time steps (i.e., lagged values) at a certain location. In Fig. 5, the autocorrelation of beam *X* is plotted against different time lags. The figure shows a strong temporal correlation in the data.

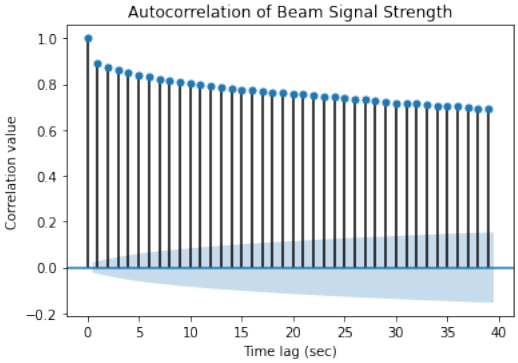

Fig. 5. Autocorrelation of wide beam *X* reported at one location

2) Spatial Analysis: the spatial correlation was obtained from the dataset, see Fig. 6. The average aggregate beam strength values were used to calculate spatial correlation. In addition, the spatial correlation is given separately for average wide beams and average narrow beams, since narrow and wide beams have different spatial exposure. For spatial correlation, the Pearson correlation coefficient is calculated between the average narrow beam strength, the average wide beam strength, the latitude and the longitude. The correlation coefficient formula is,

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} \tag{1}$$

where *cov* represents the covariance and σ is the standard deviation. From Fig. 6, it is evident that there is a strong spatial correlation between the narrow and wide beams and the latitude variable, with a correlation coefficient of 0.63 and 0.74, respectively. The beam strengths are also inversely correlated with the longitude variable, which is believed to be specific to the experiment site and setup.

The above analysis demonstrates spatial and temporal dependencies in the beam strength data, which suggests high predictability. This analysis also validates the deployment of prediction to forecast failure events and utilizes this prior knowledge to proactively prepare for an appropriate recovery mechanism.

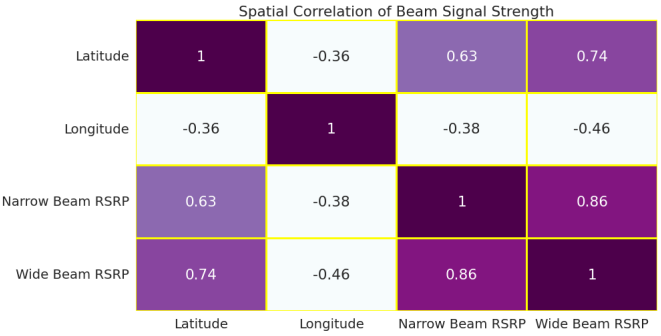

Fig. 6. Spatial correlation of beam signal strength values

IV. Simulation Results and Performance Analysis

A. Data Preprocessing

For the proposed prediction algorithms to work best, preprocessing the raw data is required. The following steps define the data preprocessing techniques applied to the dataset.

1) Data Cleaning: this study is based on real beamforming data. The data consists of beam strength location and time fingerprints. Since this study explores failures for connected-users, initial access failure is ignored. Also, its associated failure incidents were removed from the dataset by dropping all failure records reported at the beginning of the measurement session.

2) Feature Engineering: in order to make the data more informative, new features were added. The new features are summarized in Table I.

3) Up-sampling Minority Class and Down-sampling Majority Class: class imbalance problem is expected in this study because failure is an uncommon occurrence. Taking a closer look at the general statistics of the data also proves there is a class imbalance, see Fig. 7. The number of samples for the majority classes, failure-free and negligible failure, clearly exceeds the minority failure classes. Therefore, up-sampling the minority class and down-sampling the majority class are performed to ensure that the prediction algorithm has enough samples from each class to learn all classes equally.

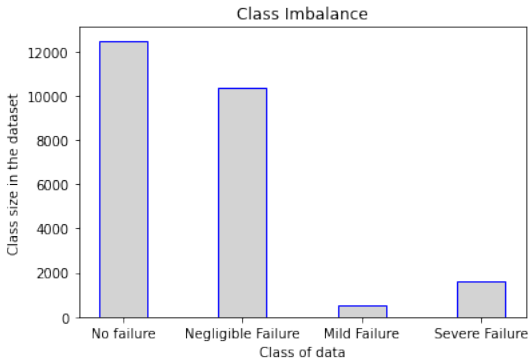

Fig. 7. Class imbalance problem in the dataset

Authorized licensed use limited to: Queen's University. Downloaded on September 18,2022 at 18:22:22 UTC from IEEE Xplore. Restrictions apply.

TABLE I

New Feature	Description
Fail flag	This feature raises a flag if there is no beam coverage (all beams measure -1).
Fail Duration	This feature measures the duration for consecutive failure incidents.
Recovery Action	Based on the duration, the suitable recovery mechanism is chosen.

B. Results

In this paper, several machine learning prediction models are tested to predict future full-beam failure and decide the appropriate recovery mechanism in case of failure. The problem is simplified to a multi-class classification problem with simple matching between the predicted failure and the appropriate recovery. To attain the true performance of each algorithm, k-fold cross-validation (or cross–k validation) is used. For k-fold cross-validation, data is shuffled and split into k groups. Each of the k groups is assigned the test group, and a model is trained on the k-1 remaining groups. The accuracies of the k tests are then averaged. Here, cross –k validation was used for k=10.

The performance of the proposed algorithms is evaluated. The prediction accuracy of the three proposed prediction models, namely, MLP, XGBoost and LSTM, is presented in Fig. 8. The prediction accuracy of the different algorithms is plotted against different data sample sizes to study the effect of the sample size on the accuracy. The figure shows that LSTM and XGBoost have a comparable accuracy of around *82%* for a small sample size, although LSTM performs slightly better than XGBoost. MLP has the worst prediction performance compared to the other two predictors, with *78%* accuracy. As the sample size goes up, MLP performance drops, and the performance of XGBoost and LSTM improves. This improvement indicates that LSTM and XGBoost are more stable with this dataset than MLP and have better generalization on larger datasets than MLP. Also, this confirms the overfitting problem in MLP. LSTM achieves the highest prediction accuracy, with *84.5%*. Hence, the rest of the analysis will be carried out using LSTM results.

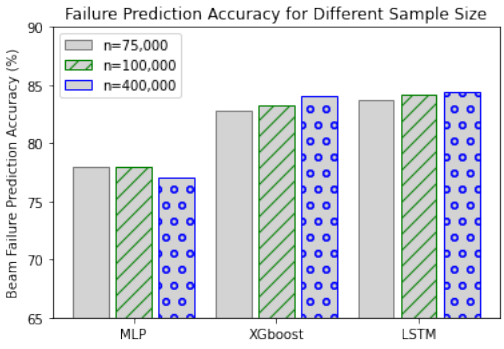

Fig. 8. Prediction accuracy of proposed algorithms

In addition to the prediction accuracy, the beam failure probability after using LSTM is calculated, i.e., the probability of unforeseen beam failure by LSTM for severe and mild failure cases. This probability is compared against the total failure probability in the dataset, the results are presented in Fig. 9. A substantial reduction in beam failure probability is apparent after using prediction. This result shows that around *0.9%* of the severe beam failure incidents were unforeseen by the prediction algorithm, whereas only *0.1%* of the mild failures were undetected by the predictor. A complimentary analysis is shown by Table II, where the beam recovery success probability of LSTM is presented. The recovery of success probability is shown for the two different classes of failure and total failure. From the table, it is shown that LSTM can successfully proactively recover *89%* of failure.

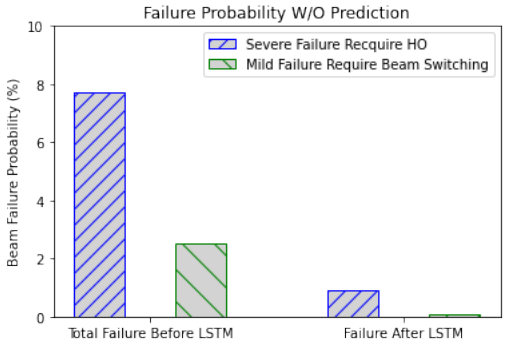

Fig. 9. The reduction in beam failure probability after using LSTM

The significance of this proposed system lies in its ability to forecast failure events, especially severe failure, and proactively prepare for recovery to avoid service discontinuity. Hence, one advantage of the proposed system is a decrease in RLF. Fig. 10 indicates this advantage by plotting the decrease in RLF, which is the gain of using prediction, against different data sample sizes. This gain is around a *75%* decrease in RLF when the model is trained on a small dataset. While the rate goes up to an *80%* decrease in RLF when the whole dataset is used for training. However, this is not the true gain of the prediction algorithm because using prediction for proactive recovery increases the probability of unnecessary HOs. HOs are expensive network functions and should be avoided, if not necessary. Therefore, this predictive beam failure and proactive recovery system comes with increased HO costs that should be taken into account when assessing the true gain from using this proposed system. Fortunately, there is only a *30%* increase in unnecessary HO as opposed to an *80%* reduction in RLF. Thus, although associated with some costs, the proposed system is efficient and can substantially improve the network performance.

Authorized licensed use limited to: Queen's University. Downloaded on September 18,2022 at 18:22:22 UTC from IEEE Xplore. Restrictions apply.

TABLE II
BEAM RECOVERY SUCCESS PROBABILITY AFTER USING LSTM

	Severe Failure	Mild Failure	Total Failure
Probability of Successful Recovery (%)	88	94	89

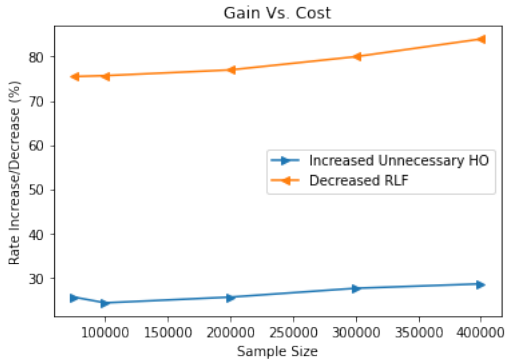

Fig. 10. Gain in decreased RLF vs. cost in increased unnecessary HO

V. CONCLUSION

To improve mmWave communication in NGCN, beam failure prediction and proactive recovery are proposed in this paper. The problems arising from passive beam failure detection and recovery are discussed and an intelligent system that forecasts future failures and acts accordingly is designed. Different ML algorithms were tested for failure prediction, the LSTM algorithm was found to outperform MLP and XGBoost in terms of prediction accuracy. Real data was used to assess the performance of the proposed system which was evaluated in terms of beam failure probability and beam recovery success rate. Finally, the true gain from using self-healing beam failure system is discussed by setting the gain in decreased RLF against the cost of increased HO rate. Because the analysis in this paper is limited to an all-beam failure, as part of future work, more complex failure scenarios will be investigated, such as a partial beam failure scenario. In addition, new network performance indicators, other than the RSRP, will be used to define the failure event, for example, the quality of service and quality of experience.

ACKNOWLEDGEMENT

The authors would like to thank Anders Berkeman, Rupesh Kumar Thakur, Per Fryking, Gustav Fahlén and Carl Drougge, from Ericsson Sweden, for administering the experiment that led to the collection of the dataset used in this paper.

REFERENCES

[1] M. Jasim and A. Aldalbahi, "Diversity coding for instantaneous link recovery in millimeter wave communications," in *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2018, pp. 7–11.

[2] F. Fernandes, C. Rom, J. Harrebek, and G. Berardinelli, "Beam management in mmwave 5g nr: an intra-cell mobility study," in *93rd Proc. IEEE Veh. Tech. Conf. (VTC2021-Spring)*, 2021, pp. 1–7.

[3] A. K. R. Chavva and N. B. Mehta, "Millimeter-wave beam selection in time-varying channels with user orientation changes," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2021.

[4] W. Ren, J. Xu, D. Li, Q. Cui, and X. Tao, "A robust inter beam handover scheme for 5g mmwave mobile communication system in hsr scenario," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2019, pp. 1–6.

[5] P. Zhou, X. Fang, Y. Fang, R. He, Y. Long, and G. Huang, "Beam management and self-healing for mmwave uav mesh networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1718–1732, 2019.

[6] Y.-N. R. Li, B. Gao, X. Zhang, and K. Huang, "Beam management in millimeter-wave communications for 5g and beyond," *IEEE Access*, vol. 8, pp. 13 282–13 293, 2020.

[7] S. Hailu, K. Pedersen, A. Karimidehkordi, F. Abinader, C. Rom, and N. Kolehmainen, "Performance analysis of mac-based beam tracking and failure detection and recovery," in *93rd Proc. IEEE Veh. Tech. Conf. (VTC2021-Spring)*, 2021, pp. 1–5.

[8] M. Mollel, M. Ozturk, M. Kisangiri, S. Kaijage, O. Onireti, M. A. Imran, and Q. H. Abbasi, "Handover management in dense networks with coverage prediction from sparse networks," in *IEEE Wireless Communications and Networking Conference Workshop (WCNCW)*, 2019, pp. 1–6.

[9] C. Jia, H. Gao, N. Chen, and Y. He, "Machine learning empowered beam management for intelligent reflecting surface assisted mmwave networks," *China Communications*, vol. 17, no. 10, pp. 100–114, 2020.

[10] W. Xu, Y. Ke, C.-H. Lee, H. Gao, Z. Feng, and P. Zhang, "Data-driven beam management with angular domain information for mmwave uav networks," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2021.

[11] N. Nasser, A. Hasswa, and H. Hassanein, "Handoffs in fourth generation heterogeneous networks," *IEEE Communications Magazine*, vol. 44, no. 10, pp. 96–103, 2006.

[12] 3GPP, "5G;NR; Requirements for support of radio resource management," 3GPP, TS 38.133, 07 2019, version 15.5.0 Release 15.

Authorized licensed use limited to: Queen's University. Downloaded on September 18,2022 at 18:22:22 UTC from IEEE Xplore. Restrictions apply.