

---

# A Decision-Tree-Based Algorithm for Proactive Handover Prediction in Multi-RAT Cellular Networks: A Drive-Test Study with Implications for 5G/6G Mobility Management

---

[Majd Hamdan](#), [Lina Yilmaz](#), [Ibraheem Shayea](#)<sup>\*</sup>, [Leila Rzayeva](#)

Posted Date: 9 May 2026

doi: 10.20944/preprints202605.0576.v1

Keywords: artificial intelligence (AI); 5G/6G mobility; machine learning; decision tree; handover prediction



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# A Decision-Tree-Based Algorithm for Proactive Handover Prediction in Multi-RAT Cellular Networks: A Drive-Test Study with Implications for 5G/6G Mobility Management

Majd Hamdan <sup>1</sup>, Lina Yilmaz <sup>1</sup>, Ibraheem Shayea <sup>1,\*</sup>  and Leila Rzayeva <sup>2</sup> 

<sup>1</sup> Department of Electronics and Communication Engineering, Istanbul Technical University.

<sup>2</sup> Research and Innovation Center "CyberTech", Astana IT University, Astana 010000, Kazakhstan

\* Correspondence: shayea@itu.edu.tr

## Abstract

The combination of ultra-dense network deployments and high mobility results in an unfavorable outcome, rendering the task of handover more difficult than in environments typical of previous generations. 5G and 6G necessitate the deployment of heterogeneous networks and small cells to meet the demand, which at the same time introduces certain challenges. This scenario introduces small cells (such as femtocells, picocells, and microcells) that have very limited coverage areas, which, combined with the high speed of user equipment, create an excessive number of handover triggers, leading to the "ping-pong effect," which wastes network resources and degrades the overall Quality of Service. Furthermore, high mobility means that a user might enter and exit a cell in less time than the mobile terminal's dwell time, dropping the connection and resulting in handover failures and radio link failures. The conventional handover methods that rely on thresholds of certain factors such as the received signal strength could be insufficient for these environments. Different criteria should be balanced to avoid the drop, such as the user's velocity, dwell time, target cell load, available bandwidth, device battery, and application latency requirements. Predictive methods could be a more efficient alternative to the existing reactive ones. This paper presents a decision-tree-based algorithm as one predictive method that learns the patterns among all the criteria mentioned and is particularly useful for avoiding ping-pongs and limiting handover failures. The classifier is trained on real multi-operator drive-test data with ping-pong events excluded from the positive class, and evaluated under Leave-One-Trace-Out cross-validation on 16 traces covering UMTS, HSUPA, HSPA+, and LTE cells. The proposed system achieves  $F_1 = 0.642$  and  $AUC = 0.797$  under LOTO, with a +0.052  $F_1$  lift over the best threshold-based baseline, while remaining interpretable and deployable in real time. The paper aims to present a solution applicable also to 5G NR and 6G.

**Keywords:** artificial intelligence (AI); 5G/6G mobility; machine learning; decision tree; handover prediction

## 1. Introduction

Cellular networks have evolved through time driven by the increasing demand for higher data rates and more seamless connectivity. 4G (Fourth Generation) has been an important transition to advanced packet-switched networks offering rates up to 100 Mbps (Ahmad, Sundararajan, Othman & Ismail, 2017). One important goal was to provide integration with other heterogeneous networking technologies (such as Wi-Fi, WiMAX, and LTE) to deliver multimedia services and seamless internet access. Moving forward, 5G has been introduced as an expansion to the bandwidth supporting enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communications (URLLC), and Massive Machine-Type Communications (mMTC) (Ahmad, Sundararajan, Othman & Ismail, 2017).

For these reasons, massive MIMO (Multiple-Input Multiple-Output) antennas and millimeter-wave (mmWave) frequencies and a dense deployment of small-cell base stations were utilized. 6G (Sixth Generation), which is a technology under academic development currently, is promising to offer higher data rates reaching up to 1 Tbps and latency as low as 0.1 ms (Loutfi et al., 2025). This technology will be AI-native, utilizing methods such as machine learning for autonomous, real-time network control (Amirova et al., 2025). Furthermore, 6G converges terrestrial infrastructures with non-terrestrial networks (NTNs), such as satellites and drones, to create an inclusive global coverage (Amirova et al., 2025). Mobility management grows to be crucial with the evolution of cellular networks. It is an important process to maintain continuous connectivity for user equipment (UEs) traversing geographical zones at different velocities. However, many struggles arise in this aspect, such as that 5G and 6G rely on high frequencies, which result in the heavy usage of small cells. This, in turn, causes the UEs passing at high velocities through the zones of coverage of those cells to trigger a high number of handovers (Amirova et al., 2025). In dense environments, poor mobility management causes many ping-pong events in which the UE rapidly and unnecessarily bounces back and forth between neighboring cells (Mohsin, Saad & Shayea, 2023). It also increases the risk of handover failures, radio link failures, packet loss, and massive signaling congestion. Effective mobility management algorithms must balance network load and execute handovers at the exact right moment. Users move across cells of different technologies (e.g., LTE, Wi-Fi, 5G...), creating heterogeneous networks (HetNet), and so the handover is called vertical. These are harder to handle than horizontal ones since comparison between criteria is harder to achieve than between similar networks (Amirova et al., 2025). Handover is considered the main method that allows mobile devices to keep a session going when moving from one place of connection to another, like between base stations or various access points, without any significant loss of service. Normally, in classic cellular networks, handover activity has three main parts—gathering information, making the decision, and finally execution or ending. In stage one, information is taken by both the UE and the network side. They gather parameters that include Received Signal Strength (RSS or RSRP), as well as measurement of signal quality (RSRQ, SINR), throughput amount, load, and some context info such as how fast the user moves, where they are, or what services they want. The second part, the decision time, is about checking if a handover should happen; then, if yes, it selects the target network that is best, using criteria around the network side, terminal, and service or user, like bandwidth, cost, user chance, battery status, coverage, latency, and personal choices. The last part, execution, means the resources will be used in the new cell, data route switched, and the exiting link will be stopped, so the move should be smooth as intended. Types of handover are horizontal (between the same Radio Access Technology (RAT), for example, LTE to LTE) or vertical, which is for different technologies, for example between LTE, WLAN, and 5G, where direct matching of signals cannot be done and so more complex selection methods are required. Poorly executed handovers can cause repeated movement between cells (called ping-pong), failing handovers, and longer waiting times, damaging Quality of Service (QoS) and Quality of Experience (QoE). Because of this, modern research sees handover as a context-aware challenge with multiple layers that need optimization, more advanced than just reaching a limit for RSS. Traditional methods for managing handovers that were deployed in 3G and early 4G do not serve useful in the modern context. Those methods relied on comparing RSS or RSRP against predefined threshold values (Elhilali, Badri & Bouami, 2023). The simplicity of those methods is indeed an advantage for computational cost but a drawback for more advanced networks. Complex, heterogeneous, and ultra-dense environments of 5G and 6G networks require different approaches since comparison becomes impossible. Criteria, along with being impossible to compare across different technologies, might actually be subject to expansion. For instance, a handover might happen to a certain cell not because it is closer or has a stronger RSS but because it offers higher bandwidth, lower monetary cost, or lower battery consumption, like Wi-Fi when finding 4G cell towers or so. Traditional handovers are governed by fixed rules known as Handover Control Parameters (HCPs) such as the Handover Margin (HOM) and the Time-to-Trigger (TTT). These parameters exist to delay the handover just long enough to ensure the signal change isn't

just a temporary fluctuation. However, using fixed HCPs creates a dangerous "Catch-22" in dense 5G networks (Mohsin, Saad & Shayea, 2023). If network operators set the threshold margin (HOM/TTT) higher to prevent unnecessary switching, the system waits too long to initiate the handover, and for high mobility scenarios, the signal might degrade quickly and cause the call to be dropped. The newer cellular networks deploy Ultra-Dense Networks (UDNs) made up of "small cells" (femtocells, picocells, and microcells) (Mohsin, Saad & Shayea, 2023). The dwell time, which is the duration of a UE in a certain cell's coverage, isn't accounted for in traditional methods. They will initiate a handover to a small cell simply because the signal is strong, ignoring that the user will exit that cell a second later (Goh et al., 2023), wasting a considerable amount of processing power caused by the unnecessary handover. In ultra-dense environments, traditional algorithms trigger handovers constantly, which creates an increase in signaling traffic that burdens the core network and heavily consumes the processing power of the base stations (Xenakis, Passas, Merakos & Verikoukis, 2014). A break-before-make (Gupta et al., 2021) process is followed when making a hard handover, breaking the old radio connection before reaching the new one. Packets are buffered and delayed during this switch. If handovers happen too frequently due to bad threshold logic, the accumulated delay and packet loss make real-time applications (like autonomous driving, VoIP, or live video streaming) impossible to sustain (Gupta et al., 2021). To overcome these issues, this paper proposes a Machine Learning model that works to proactively predict a handover by learning the pattern among the different features of networks rather than focusing on thresholds. The paper examines the efficiency of decision trees in separating and classifying features to predict handovers, avoiding the ping-pong effect. A dataset recorded between late 2017 and January 2018 has been captured that includes the movement of a vehicle between cells of different technologies. The purpose of the paper is to develop an efficient algorithm, explore its strengths and weaknesses, and set a direction for future research.

**Table 1.** Dataset and Training Configuration.

Parameter	Value
Trace files (Operator A / B)	16 (9 / 7)
Total time-series rows	10,783
Cell-change events detected	211 (206 with valid prediction window)
Positive samples ( <i>Pending_Handover</i> )	167 (after ping-pong exclusion)
Negative samples ( <i>No_Handover</i> , 2:1)	334
Total training samples	501
Number of features	17 (13 engineered + 4 NRx neighbor-cell)
Prediction window	1–3 s before cell change

## 2. Related Work

In cellular networks, the handover decision is an important process that helps keep a stable connection when a mobile user moves between different cells. Traditional handover methods in early cellular systems mainly depend on mathematical and rule-based techniques that use radio signal measurements to trigger a handover. For example, indicators such as Received Signal Strength (RSS) were commonly used to decide when a user should switch to another cell (Elhilali, Badri & Bouami, 2023). In more advanced cellular systems, other measurements such as Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), and Signal-to-Interference-plus-Noise Ratio (SINR) are also used to measure link quality (Ahmad, Sundararajan, Othman & Ismail, 2017). These methods usually use predefined thresholds to decide when the signal of the serving cell becomes too weak. To reduce unnecessary handovers and avoid the ping-pong effect, techniques such as hysteresis margins and Time-to-Trigger (TTT) are often used in traditional mobility management schemes (Xenakis, Passas, Merakos & Verikoukis, 2014). However, these approaches mainly depend on signal strength measurements and may not work well in dense and highly dynamic network environments (Mohsin, Saad & Shayea, 2023). In recent years, many researchers have explored the use

of machine learning techniques to improve mobility management and handover decisions in cellular networks. Unlike traditional rule-based methods, machine learning models can analyze several network parameters together and learn useful patterns from previous data (Loutfi et al., 2025). This helps the network make better decisions based on different conditions instead of depending only on signal strength. Some studies also show that machine learning can predict handover events with higher accuracy by using different features such as signal quality, user speed, and mobility behavior (Mohsin, Saad & Shayea, 2023). Because of this, intelligent handover management has become an important research area for next-generation wireless networks where network conditions can change very quickly (Elhilali, Badri & Bouami, 2023). Among the different machine learning models, the Decision Tree algorithm is a suitable and practical choice for this study. Decision Trees create a clear and simple structure that makes it easier to see how different input features affect the final decision (Breiman, Friedman, Olshen & Stone, 1984). Because of this structure, researchers can better understand how different network conditions influence the handover decision process. Compared to more complex models such as deep neural networks, Decision Trees usually need less computational power and are easier to use in real-time network systems (Mohsin, Saad & Shayea, 2023). Decision Trees are also simple to understand, which is one reason why they are widely used for classification problems that involve structured network data (Elhilali, Badri & Bouami, 2023). For this reason, in this study we use a Decision Tree model to predict handover events using real network measurement data.

### 3. Methodology

#### 3.1. Network Environment

The proposed system is tested using real cellular network traces collected from two different mobile operators, called Operator A and Operator B in this study. The dataset represents a heterogeneous multi-RAT (Radio Access Technology) environment where the User Equipment (UE) moves between different network technologies during normal mobility. The network environment includes several radio technologies that appear in the dataset through the NetworkMode parameter. These technologies include UMTS, HSUPA, HSPA+, and LTE. To make the processing easier inside the machine learning pipeline, the different network modes are converted into numerical values. This allows the algorithm to recognize the different radio technologies and detect both intra-RAT handovers (inside the same technology) and inter-RAT handovers (between different technologies). The dataset reflects realistic user movement. The traces show continuous UE movement as the device connects to different base stations over time. Each base station has its own CellID. When this value changes, it means the device has switched to another serving cell. The user movement can also be seen from the recorded UE speed and from the change in distance between the UE and the serving cell. The goal of the proposed system is to analyze the cellular network measurements collected from the user equipment and predict if a handover may happen in the near future.

#### 3.2. Dataset Description

In this study, we use a dataset that contains 16 network trace files stored in CSV format. Each file was recorded during a different measurement session where a mobile device moved through the cellular network coverage area and recorded the network conditions. The traces include time-ordered measurements such as signal strength indicators, signal quality values, network identifiers, and mobility information like speed and location. These measurements help us understand the network conditions that the mobile device experiences while it moves through the coverage area. Our dataset represents time-series cellular measurements collected during drive-test experiments. In these experiments, the user equipment moves through different network coverage areas while the network parameters are recorded. Each record shows the network conditions that the mobile device observes at a specific moment. The dataset also includes measurements from different radio technologies such as UMTS, HSUPA, HSPA+, and LTE. This means that during the measurements the mobile device can move between different types of cellular networks. Because of this, we can observe different

network behaviors and different signal conditions. In total, our dataset contains 16 measurement traces collected during different mobility sessions. These traces show realistic user movement and different changes in network conditions over time. This helps us study how network conditions change when the user moves through the cellular network.

The 16 CSV traces were recorded between 30 November 2017 and 27 January 2018 with two operators, Operator A and Operator B, on a bus trip. Each row is a one-second sample of measurement of mobility features. The dataset has been released by Raca *et al.* 2018 (Raca *et al.*, 2018) and their paper includes details about methodology and collection. No human-subjects data were involved and only radio-interface KPIs and cell identifiers were used.

### 3.3. Applicability to 5G NR and 6G

The paper suggests in its abstract and title the applicability to 5G and 6G. Therefore, it is worth mentioning how this applicability could be realized. First, the paper identifies the elements that could be transferred from a 3G/4G decision-tree system to a 5G and 6G decision-tree system. The pre-handover window feature could be transferred since its extraction is radio-technology-agnostic and applies to NR SSB-RSRP, SSB-RSRQ, and L1-SINR measurements. Second, the main contribution of this paper — the LOTO implementation — could also be transferred since it is a protocol-level concern and does not depend on the RAT. The training-time ping-pong exclusion could also be incorporated if the oscillation thresholds are adjusted to the shorter NR cell dwell times. The NRx neighbor-cell block has direct NR analogues (best-beam SSB-RSRP, second-best-beam SSB-RSRP, inter-beam margin) and therefore transfers with minor modification. However, other elements cannot be seamlessly transferred: NR Conditional Handover (3GPP, 2023a) introduces a prepare-early, execute-on-condition semantics that the single-shot 1–3 s pre-handover window prediction cannot capture; NR beam-level mobility creates a new intra-cell event class that does not appear in the CellID-change labeling; and Multi-RAT Dual Connectivity (MR-DC/EN-DC) makes the binary *Pending\_Handover* vs. *No\_Handover* label scheme less relevant due to the secondary connection and the need for additional classes. A natural next step is to evaluate the present pipeline on a public 5G NR dataset and report the incompatibilities with the current system together with the changes needed to make it compatible.

### 3.4. Data Preprocessing and Data Generation

Before training the machine learning model, several preprocessing steps are applied to organize the raw network traces and prepare the dataset. The raw measurements are stored in multiple CSV files. During preprocessing, these files are loaded and processed one by one. The timestamps are parsed and the samples are arranged in chronological order so that the data follows the real timeline of the measurements. In this way, each record represents the network state observed by the user equipment at a specific moment. Next, handover events are identified in the dataset. In this study, a handover (HO) event is defined as the moment when the CellID changes between two consecutive samples. This change indicates that the user device moves from one serving cell to another cell in the network. After detecting the handover events, samples are extracted from a prediction window before each handover occurs. We search for records that appear between 1 and 3 seconds before the handover time and select the sample closest to 2 seconds before the event. These records represent the network conditions shortly before the handover happens. This time window is selected because network conditions usually start to change shortly before a handover occurs, which makes this period useful for prediction. We also collect samples that represent normal network behavior. These samples are taken from periods where the serving cell does not change. In other words, the device stays connected to the same cell during this time. To make sure that the network is in a stable condition, we only select samples that are at least 10 seconds away from any handover event. This helps us represent normal network behavior without the effect of nearby handovers. After completing these preprocessing steps, we obtain a structured dataset that contains two types of samples: pre-handover samples and stable network samples. The resulting dataset is then used to train and evaluate the proposed machine learning model in the following stages. Ping-pong events that are cell changes

satisfying 3GPP TR 36.839 ping-pong criteria are excluded from the positive class before training to prevent the classifier from learning them and mistaking them for genuine handovers. Out of the 206 cell-change events with a valid 1–3 s prediction window, 39 are ping-pong events and are removed, leaving 167 positive samples (*Pending\_Handover*). Negatives are undersampled at a 2:1 ratio, giving 334 *No\_Handover* samples and a total of 501 training samples.

### 3.5. Feature Engineering

To help the machine learning model understand the network conditions before handover events, we extract 13 base features from the measurements recorded in the network traces (*RSRP*, *RSRQ*, *SNR*, *CQI*, *RSSI*,  $\Delta RSRP$ ,  $\Delta RSRQ$ ,  $RSRP_{std}$ , *Speed*, *ServingCell\_Distance*, *ServingCell\_Distance\_Delta*, *NetworkMode*, and *TimeSinceLastHO*); an additional 4 neighbor-cell (NRx) features are introduced in Section 3.6.1, bringing the total used by the model to 17. These features describe the radio signal quality, how the signal changes over time, and the mobility state of the user. In this work, the system builds a supervised dataset for handover prediction. The Decision Tree classifier uses both the direct signal measurements and the short-term signal behavior that appears shortly before a handover happens. For easier explanation, the input features used by the model are divided into several groups.

**Signal Strength and Quality:** These features describe the radio link between the user equipment (UE) and the serving cell. They include *RSRP* (Reference Signal Received Power), *RSRQ* (Reference Signal Received Quality), *SNR* (Signal-to-Noise Ratio), *CQI* (Channel Quality Indicator), and *RSSI* (Received Signal Strength Indicator). These measurements give information about the signal strength, the level of interference, and the overall quality of the wireless connection.

**Signal Dynamics:** Besides the direct signal values, we also consider how the signal changes over time. In particular, the changes in *RSRP* and *RSRQ* are calculated using the difference between consecutive samples. These features allow the model to notice fast signal degradation, which often appears shortly before a handover event. To further describe signal behavior, we also compute the standard deviation of the *RSRP* values over a short window of five samples. This value indicates how much the signal varies during a short period of time. Such variations may occur because of fading, shadowing, or when the user is close to the edge of a cell. The change in signal strength is computed using the first-order difference between consecutive samples:

$$\Delta RSRP(t) = RSRP(t) - RSRP(t - 1) \quad (1)$$

$$\Delta RSRQ(t) = RSRQ(t) - RSRQ(t - 1) \quad (2)$$

These features help the model detect fast signal degradation that often happens shortly before a handover event. To measure signal instability, the standard deviation of *RSRP* is calculated over a short lookback window of five samples:

$$RSRP_{std} = \text{Std}(RSRP_{t-5}, \dots, RSRP_t) \quad (3)$$

**Mobility and Location Features:** User movement is an important factor in handover decisions. Because of that, we include features that describe how the UE moves relative to the serving base station. These features include the user speed recorded in the trace, the distance between the UE and the serving cell (*ServingCell Distance*), and how this distance changes over time (*ServingCell Distance Delta*). Using these values, the model can better understand whether the user is moving away from the serving cell, which is a situation that often leads to a handover. The distance change is computed as:

$$\Delta d(t) = d(t) - d(t - 1) \quad (4)$$

**Network Context Features:** Besides signal and mobility information, we also include features that describe the current network situation. One of these features is *NetworkMode*, which indicates the radio technology currently used by the device, such as UMTS, HSUPA, HSPA+, or LTE. Another

feature is Time Since Last Handover, which measures how much time has passed since the previous cell change. Together, these features give additional context about the UE connection and the network that is serving it.

**Data Cleaning:** Before the model is trained, the dataset is checked for missing values. In some network traces, certain parameters like SNR, CQI, or RSSI are not always available. When this happens, the missing values are replaced with zeros. This simple step allows the model to process the dataset without errors even when some measurements are missing.

**Neighbour-Cell (NRx) Features:** To let the model be proactive in choosing the target cell, a neighbor-cell block is added to the feature vector. At each sample the neighbor cells are ranked by RSRP: (i) the best-neighbor RSRP, (ii) the second-best-neighbor RSRP, (iii) the neighbor margin, defined as  $\Delta_{nbr} = RSRP_{best\_nbr} - RSRP_{serving}$ , and (iv) the count of neighbor cells whose RSRP exceeds the detectability threshold. This combination lets the tree make its prediction based on the possibility of a suitable target cell existing. The NRx features give the tree more awareness and are not strongly correlated with the other features.

**Positive-Class Definition:** Ping-pong events are removed from the positive class at training time so the classifier learns to predict only genuine handovers directly from the pre-handover window.

### 3.6. Machine Learning Model

The handover prediction system that this paper proposes uses a Classification and Regression Tree (CART), also referred to as a decision tree, which is a supervised learning algorithm. Two outcomes are designed: *Pending\_Handover* or *No\_Handover*, mapped to a vector of features that include some mobility criteria and some radio measurement ones. The decision tree splits are highly related to a meaningful threshold of radio parameters that help domain experts to investigate the logic (Breiman, Friedman, Olshen & Stone, 1984).

#### 3.6.1. Feature Space

Along with the features already existing in the dataset, 5 new features have been obtained from the already existing ones to enable the model to better detect patterns. Notably, a pre-handover window covering the 1–3 rows (at the  $\sim 1$  Hz sampling rate) preceding each cell-ID change was engineered. The window allows for proactive detection since it explores the tendencies of features right before a handover happens. Table 2 presents the features in four groups.

**Table 2.** Feature Groups Used in the Decision Tree Classifier.

Category	Features	Description
Instantaneous radio	RSRP, RSRQ, SNR, CQI, RSSI	Signal strength, quality, interference at measurement point
Temporal dynamics	RSRP_delta, RSRQ_delta, RSRP_std_before	First-order differences and windowed standard deviation capturing degradation trends
Mobility context	Speed, ServingCell_Distance, ServingCell_Distance_delta	UE velocity and change in distance to the serving cell
Network / history	NetworkMode, time_since_last_HO	RAT encoding (UMTS= 1, HSUPA= 2, HSPA+= 3, LTE= 4) and time elapsed since the previous cell change

*RSRP\_delta*, *RSRQ\_delta* represent the rate of change in RSRP and RSRQ, respectively, between their consecutive measurements. *RSRP\_std\_before* represents the standard deviation of RSRP over a window of 5 samples to quantify signal instability. *time\_since\_last\_HO* represents the time since the last change of cell to reflect the clustering of handovers happening at the edge of cells.

#### 3.6.2. Classification Algorithm

The CART algorithm chooses the best split by minimizing the Gini impurity through the selection of proper features and thresholds.

$$G(t) = 1 - \sum_i p_i(t)^2 \quad (5)$$

where  $p_i(t)$  is the proportion of samples belonging to class  $i$  at node  $t$ . Partitioning continues until the stopping criteria ( $\text{MinLeafSize} \geq 10$ ,  $\text{MaxNumSplits} \leq 30$ ) are satisfied. Each leaf is assigned the majority class of the training samples that reach it. The trained tree has 18 split nodes and 19 leaf nodes. The tree begins by splitting on  $\text{NRxRSRP}_{\text{missing}} < 0.5$  to partition samples by neighbor-cell measurement availability, and the following nodes test the following measurements:  $\text{NRxRSRQ}$ ,  $\text{RSRQ}$ ,  $\text{RSRP}$ ,  $\text{CQI}$ ,  $\text{RSRP}_{\text{std\_before}}$ ,  $\text{RSRP}_{\text{delta}}$ ,  $\text{RSRQ}_{\text{delta}}$ ,  $\text{SNR}$ ,  $\text{RSSI}$ ,  $\text{Speed}$ , and  $\text{ServingCell\_Distance}$ , which are radio and mobility measurements indicating the state of signal before a cell change, creating the proactiveness desired.

The hyperparameters in Table 3 specify the tree configuration.

**Table 3.** Decision Tree Hyperparameters.

Parameter	Value	Justification
<i>MaxNumSplits</i>	30	Upper bound on tree complexity; trained tree used 18 splits
<i>MinLeafSize</i>	10	Each leaf represents $\geq 10$ samples, preventing overfitting
<i>Split criterion</i>	Gini	Standard impurity measure for CART classification
<i>NaN imputation</i>	0	Out-of-range sentinel for negative-valued radio features
<i>Random seed</i>	42	Reproducibility of undersampling

Ping-pong events are identified using two rules when preparing the data and then labeled for removal from the positive class before training begins. Rule one is the dwell-time rule: when the user equipment returns to the previous serving cell within a dwell-time threshold, the event is flagged as a ping-pong; separate thresholds are applied for intra-RAT and inter-RAT cases. Rule two is the oscillation rule: when three or more cell changes occur within a 3-second observation window, the fast back-and-forth switching indicates a ping-pong effect. These rules and thresholds follow the 3GPP TR 36.839 recommendations (3GPP, 2013). The specific values  $\tau = 5 \text{ s} / 3 \text{ s}$  and the “ $\geq 3$ -cell-changes-in- $\pm 3 \text{ s}$  among  $\leq 2$  cells” oscillation rule are this work’s operational thresholds inspired by TR 36.839.

### 3.7. Deployment Surface

The classifier is designed to be deployed as an xApp on the O-RAN Near-RT RAN Intelligent Controller with a control loop of 10 ms to 1 s, restricted by the feature pipeline and not the model, since it is light and fast enough. The trained decision tree has 18 splits and evaluates in sub-microsecond time on a single-core ARM device, but the  $\text{RSRP}_{\text{std\_before}}$  feature requires a 5-sample backward window at 1 Hz, introducing an inherent  $\sim 5 \text{ s}$  feature-assembly latency. The tree is therefore placed downstream of the Near-RT RIC’s KPM (Key Performance Measurement) consumer, which ingests  $\text{RSRP}$ ,  $\text{RSRQ}$ ,  $\text{SNR}$ ,  $\text{CQI}$ ,  $\text{RSSI}$ ,  $\text{speed}$ , and neighbor-cell  $\text{RSRPs}$  over the E2-SM-KPM interface.

At inference time, the classifier emits a *Pending\_Handover* signal 1 to 3 seconds before the predicted cell change. The operator-side policy layer translates the signal into a proactive measurement configuration update delivered over the A1 interface, and it lowers the target neighbor’s TTT or raises the HOM so that the next 3GPP A3 event fires earlier and finds a better target choice. The classifier pre-conditions the standardized A3 event-triggered handover execution path; if the prediction is wrong, the A3 fall-back remains in force. The simple model trades marginal accuracy for auditability and compatibility with existing 3GPP-compliant RRC mobility procedures (3GPP, 2023b). Other researchers (Tayyab, Gelabert & Jäntti, 2019) survey the transition from LTE-style handover signaling to NR mobility and dual-connectivity-based handover optimization for 5G mmWave environments (Polesse et al., 2017).

### 3.8. Training and Validation Procedure

#### 3.8.1. Dataset

The dataset includes 16 CSV files from two mobile network operators, A and B, over overlapping routes. The recordings are from 30 November 2017 to 27 January 2018, totaling 10,783 time-stamped measurement rows of one second each. Therefore, 1-second changes are often row changes. Each row records RSRP, RSRQ, SNR, CQI, RSSI, UE coordinates, speed, serving cell identity, serving cell coordinates, and network mode. The dataset contains raw 211 cell ID changes, not taking into consideration any ping-pong effects or so. Of these, 206 had at least one valid measurement within the 1–3 s prediction window and were retained as positive-class samples (*Pending\_Handover*). The remaining 5 events occurred too rapidly for any measurement to fall within the window and were excluded. Of those 206 valid-window events, 39 are identified as ping-pong (per the 3GPP TR 36.839 criteria) and removed from the positive class, leaving 167 genuine-handover positives for training.

#### 3.8.2. Negative-Class Sampling and Balancing

A severe imbalance exists between the negative class (*No\_Handover*) and the positive class (*Pending\_Handover*), and therefore, an undersampling has been conducted to ensure a ratio of 2:1. With 167 genuine-handover positives (after ping-pong exclusion), this yields 334 *No\_Handover* samples and a total training set of 501 samples. Samples at least 10 seconds away from cell-change events were selected to ensure similarity and avoid biases in undersampling.

#### 3.8.3. Validation Strategy

Temporal data leakage was a key concern when evaluating the classifier. Standard k-fold cross-validation was considered but rejected, as it randomly assigns samples to folds, allowing temporally close measurements from the same drive test to appear in both training and test sets. This would flaw the model since it would be tested on highly correlated data (near-unity autocorrelation), inflating performance and preventing it from generalizing to unseen data. For this reason, a Leave-One-Trace-Out (LOTO) cross-validation scheme was employed instead. In every fold, all samples from one trace are held out as the test set, and the remaining 15 traces form the training set. A new decision tree is trained per fold, and predictions are accumulated across all folds to compute aggregate metrics. This method confirms the model's ability to generalize to data not included in the training set (Kohavi, 1995).

#### 3.8.4. Hyperparameter Configuration

Table 3 shows the hyperparameters used for the CART classifier. *MaxNumSplits* was chosen after trial and error to be 30, providing good capacity. The tree has 18 splits, indicating this limit was non-binding. *MinLeafSize* was chosen to be 10 to avoid overfitting, especially after downsampling was performed. For missing RSRP and RSRQ values, dead-signal sentinels of  $-140$  dBm and  $-20$  dB were imputed; for the remaining features (SNR, CQI, RSSI), missing values were imputed with 0 as an out-of-range sentinel. A random seed of 42 was chosen to ensure reproducibility of the negative-class undersampling.

LOTO's external validity depends on strict separation between training and test folds; every preprocessing step used in this study is explicitly verified to operate within-trace only. (a) All missing-value sentinels ( $-140$  dBm for RSRP,  $-20$  dB for RSRQ, 0 for the remaining features) are fixed constants that are not fit on data, so they carry no cross-fold information. (b) The windowed feature *RSRP\_std\_before*, the first-order difference features ( $\Delta$ RSRP,  $\Delta$ RSRQ,  $\Delta d$ ), and *time\_since\_last\_HO* are computed per trace, with NaN at trace boundaries; no window crosses a trace boundary. (c) The ping-pong labeling rules (Section 3) are applied per trace using only cell-change events from within the same trace, never peeking into samples from another trace. (d) The CART hyperparameters *MaxNumSplits*=30, *MinLeafSize*=10, and the Gini split criterion are fixed a priori on methodological grounds (capacity control, overfitting protection) and are not tuned against the LOTO aggregate metric.

Under these conditions, no feature or label used at inference on the held-out trace was observed at training time.

#### 4. Key Performance Indicators (KPI)

Key Performance Indicators (KPIs) are metrics used to evaluate system performance from the user's perspective. They are crucial since the user side is key to providing good user satisfaction and tackling the issues that do not show up in the infrastructure-level metrics. KPIs measure the accuracy of the prediction mechanism and the quality of the resulting handover decisions. The following sections explain each KPI used in this research, including how it is computed and how it impacts network performance.

##### 4.1. Handover Probability (HOP)

Handover Probability (HOP) measures the probability that a UE will go through a handover process. It is a main metric in analyzing the mobility behavior of a network. HOP is defined as the ratio of handover events to the total number of measurement samples:

$$\text{HOP} = \frac{N_{HO}}{N_{total}} \quad (6)$$

where  $N_{HO}$  is the number of detected handover events and  $N_{total}$  is the total number of measurement rows on all traces. The significance of HOP's lies in its measurement of the mobility intensity in the network. An increase in HOP indicates an increase in the probability that a handover has occurred, increasing the signaling load on the core network; this requires more consumption of radio resources and increases the chance of radio failures as well. Overall, the objective of an optimized network is to decrease HOP while maintaining seamless connectivity. Because ping-pong events are excluded from the positive class at training time, the effective HOP corresponds to the handover rate when only genuine (non-ping-pong) handovers are counted:

$$\text{HOP}_{\text{effective}} = \frac{\text{Detected HOs} - \text{Ping-Pongs}}{N_{total}} \quad (7)$$

where *Detected HOs* are cell-ID change events, *Ping-Pongs* are the cell-change events labeled as ping-pong, and  $N_{total}$  is the total number of measurement rows in all traces.

##### 4.2. Handover Ping-Pong Probability (HOPP)

Handover Ping-Pong Probability (HOPP) measures the percentage of ping-pong handovers in a system. HOPP is defined as the ratio of ping-pong handover events to total handovers:

$$\text{HOPP} = \frac{N_{\text{PingPong}}}{N_{HO}} \quad (8)$$

where  $N_{\text{PingPong}}$  is the number of handovers that meet the ping-pong criteria (short dwell time, return to previous cell, or rapid oscillation) and  $N_{HO}$  is the total number of handover events. Ping-pong handover events are a major drawback in a network since they cause double signaling overhead and might result in potential data interruption, decreasing the Quality of Service (QoS), as well as being a waste of radio resources. An increase in HOPP indicates that the network is making too many unnecessary handover decisions, often caused by signal fluctuations at cell edges (shadow fading), inappropriate hysteresis/TTT settings, or a lack of predictive intelligence. HOPP is a dataset-level property (18.93% across the 16 traces). The effective HOPP corresponds to the classifier's false-positive behavior on ping-pong windows, i.e., how often a cell-change event that is actually a ping-pong is still predicted as a genuine handover:

$$\text{HOPP}_{\text{eff}} = \frac{\text{Ping-Pongs predicted as genuine}}{\text{Total actionable HOs}} \quad (9)$$

where the numerator is obtained directly from the LOTO confusion matrix on held-out ping-pong events, and the denominator is the number of actionable handovers emitted by the classifier. Because the tree is trained without ping-pong positives, this quantity is naturally driven toward zero in expectation; the empirical value on held-out ping-pong windows is reported in the accompanying code artefact.

#### 4.3. Radio Link Failure (RLF)

Radio Link Failure (RLF) occurs when the radio connection is lost. In 3GPP specifications, RLF is defined as when RSRP drops below a defined threshold ( $Q_{out}$ ) and remains below it for a timer duration ( $T_{310}$ ), or when the UE fails to successfully complete a handover within the allowed time. The RLF rate is defined as:

$$\text{RLF Rate} = \frac{N_{RLF}}{N_{HO \text{ attempts}}} \quad (10)$$

Where  $N_{RLF}$  is the number of RLF and  $N_{HO \text{ attempts}}$  is the number of handover attempts. When an RLF occurs, the UE must go through a connection re-establishment procedure, which starts by detecting the failure ( $T_{310}$  timer expiry), then scanning for a suitable cell, then performing Random Access on the new cell, and finally re-establishing the RRC connection. This entire process can take several hundred milliseconds to several seconds, during which the user experiences service outage. For real-time services such as VoLTE or video calls, even a single RLF can cause noticeable degradation. RLFs in the context of handover can be categorized as: Too-Late Handover (the serving cell signal drops below the usable threshold before the handover completes), Too-Early Handover (the handover is triggered too early when the target cell is not yet ready or the UE is moving away from the target), and Handover to Wrong Cell (the handover targets a cell that is not the optimal one, leading to immediate degradation).

#### 4.4. Data Rate

Data rate is defined as the throughput expressed in kilobits per second (kbps). In the dataset used in this research, it is split into downlink (DL) and uplink (UL). Each sample, one row or one second, has its own DL and UL. The behavior of the data rate is very crucial for the user experience. It is directly linked to the QoE and therefore remains important to maintain user satisfaction. While the data rate must remain high, or as promised commercially to the user, it might undergo some drawbacks during a handover. Firstly, a brief interruption of the session might happen, forcing the data rate to drop to 0. Secondly, after a handover, the UE might try to rebuild the TCP/IP connections, slowing

Table 4. KPI Summary.

KPI	Formula	Goal	System Impact
HOP	$N_{HO} / N_{total}$	Minimize	Reduced because ping-pongs are not predicted by the classifier.
HOPP	$N_{PP} / N_{HO}$	Minimize toward 0	Minimized by design; classifier trained with ping-pong events excluded from the positive class (empirical HOPP <sub>eff</sub> reported separately in the code artefact).
RLF	$N_{RLF} / N_{HO \text{ attempts}}$	Minimize	Early prediction reduces too-late HOs
Data Rate	DL/UL throughput (kbps)	Maximize	Fewer HOs result in fewer throughput dips

down the data rate; finally, the scheduler at the new cell might need time to learn the UE's channel conditions and therefore allocate optimal resources. Ping-pongs significantly contribute to downgrading the data rate, as the data rate might be dropped twice between the cells. The proposed classifier is trained not to predict ping-pongs, so the associated throughput dips are avoided at the prediction stage, maintaining more stable data rates during mobility. The data rate is analyzed by comparing the DL and UL in three conditions: (a) during stable periods when there is no handover in

the surrounding 10 seconds, which serves as the baseline data rate, (b) in the 5-second window around actual handover events, showing the impact of handovers on throughput, and (c) in the 5-second window around ping-pong events specifically, showing the additional damage caused by unnecessary handovers. The difference between (a) and (c) shows the data rate penalty attributable to ping-pong handovers that the system aims to eliminate.

## 5. Results Analysis & Discussion

### 5.1. Metrics Used

The performance was measured through the classical metrics: Accuracy, Precision, Recall, and F1-Score. A confusion matrix is obtained that includes TP, FP, FN, and TN, denoting the counts of true positives, false positives, false negatives, and true negatives, respectively.

**Accuracy** measures the overall fraction of correctly classified samples:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

**Precision** measures the fraction of predicted handovers that are genuine. A high precision indicates few false alarms, which is important in network management where unnecessary handover preparation wastes signalling resources:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (12)$$

**Recall** measures the fraction of actual handovers that the model successfully detects. A low recall means the system misses genuine handovers, potentially causing radio link failures:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (13)$$

**F1-Score** is the harmonic mean of precision and recall, providing a single balanced metric when both false positives and false negatives carry a cost:

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

In addition, the Receiver Operating Characteristic (ROC) curve and its area under the curve (AUC) were computed to evaluate classification performance across all possible decision thresholds, providing a threshold-independent measure of discriminative power.

### 5.2. Results Analysis

This section presents the results of the proposed system. The pipeline is a single Decision Tree trained to predict genuine (non-ping-pong) handovers directly. Results are reported under Leave-One-Trace-Out (LOTO) cross-validation using the metrics and KPIs defined above, together with baselines (no-neighbor DT, RSRP threshold, RSRP<sub>Δ</sub> threshold, A3-like NRx margin) for comparison.

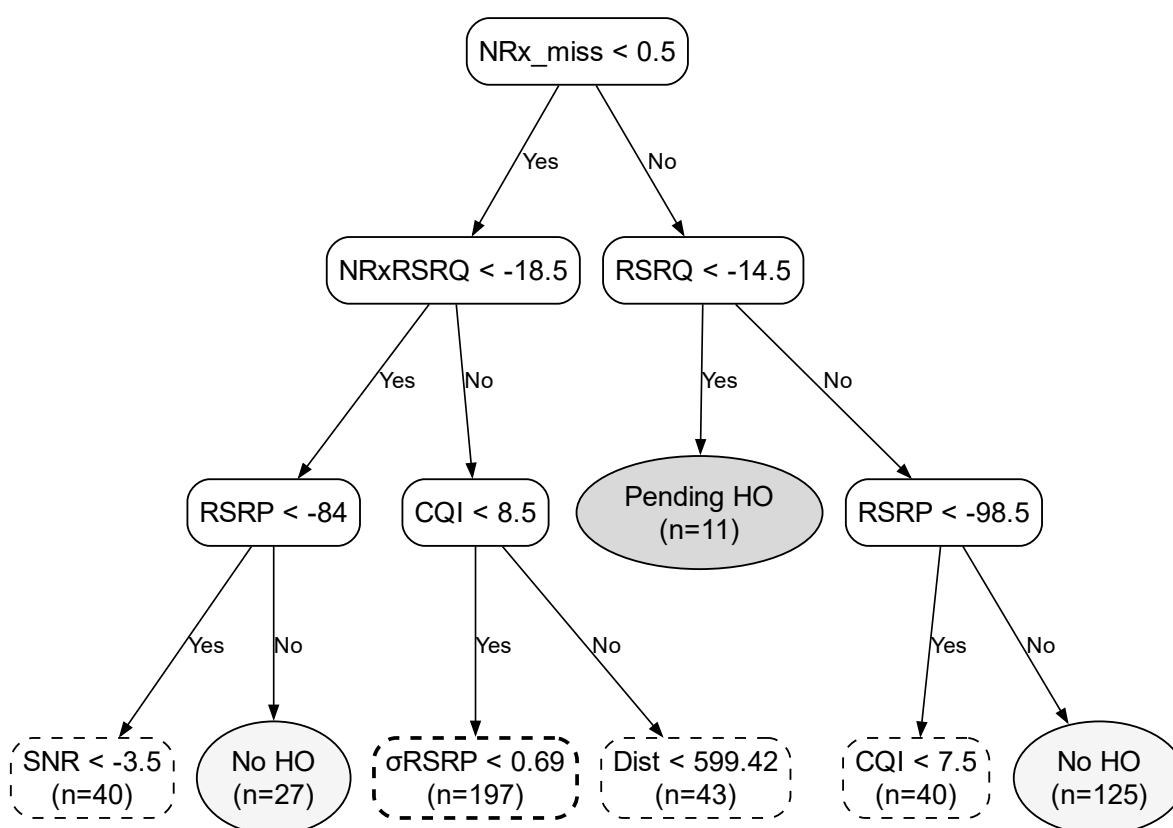
#### 5.2.1. Handover Prediction

Table 5 presents the aggregate classification results across all 16 LOTO folds. The model achieves an overall accuracy of 76.8% and an AUC of 0.797, showing good discriminative ability between pre-handover and stable radio conditions. The precision of 0.662 shows that there is an acceptable amount of false alarms. False alarms are disadvantageous for resources since they consume a considerable amount of it. Recall of 0.623 shows that approximately 62% of actual handovers are detected in advance; the 63 missed detections (false negatives) include cases where signal conditions degraded too rapidly for the pre-handover features to exhibit clear signatures within the 1–3 s window. The F1-score of 0.642 reflects a reasonable balance between these two objectives. The precision–recall tradeoff favors

precision, which is appropriate for proactive network management: falsely predicting a handover (and allocating resources on a target cell) is more wasteful than occasionally missing a prediction, since the standard A3 event-triggered mechanism remains as a fallback. The AUC of 0.797 shows that the classifier discriminates well across the operating points. The ROC curve shown in Figure 4 rises sharply at low false-positive rates, confirming the model achieves high sensitivity in the regime where false alarms must be minimized. A comparison against baseline methods is provided after Table 6.

**Table 5.** Handover prediction results (LOTO).

Metric	Value
Accuracy	76.8%
Precision (Handover)	0.662
Recall (Handover)	0.623
F1-Score	0.642
AUC	0.797
Specificity	0.841
NPV	0.817
True Positives / False Negatives	104 / 63
False Positives / True Negatives	53 / 281



Yes = condition true (left); No = condition false (right). Dashed boxes denote subtrees expanded in Table 6.

**Figure 1.** Pruned decision tree showing the top three levels. Abbreviated labels:  $NRx\_miss = NRxRSRP\_missing$ ,  $\sigma RSRP = RSRP\_std\_before$ ,  $Dist = ServingCell\_Distance$ .

**Table 6.** Complete Decision Tree Node Descriptions (18 Split Nodes, 19 Leaf Nodes). Leaf rows report the assigned class together with the training-sample count  $n$  and the within-leaf class breakdown (P = Pending\_HO, NH = No\_HO).

Node	Split Condition	Parent	Yes →	No →	Leaf Class (n: P/NH)
1	$NRxRSRP\_missing < 0.5$	Root	Node 2	Node 3	—
2	$NRxRSRQ < -18.5$	Node 1 (Y)	Node 4	Node 5	—
3	$RSRQ < -14.5$	Node 1 (N)	Leaf	Node 7	—
4	$RSRP < -84$	Node 2 (Y)	Node 8	Leaf	—
5	$CQI < 8.5$	Node 2 (N)	Node 10	Node 11	—
6	— (leaf)	Node 3 (Y)	—	—	Pending_HO (n = 11: 7P/4NH)
7	$RSRP < -98.5$	Node 3 (N)	Node 12	Leaf	—
8	$SNR < -3.5$	Node 4 (Y)	Leaf	Leaf	—
9	— (leaf)	Node 4 (N)	—	—	No_HO (n = 27: 0P/27NH)
10	$RSRP\_std\_before < 0.69$	Node 5 (Y)	Node 16	Node 17	—
11	$ServingCell\_Distance < 599.42$	Node 5 (N)	Leaf	Leaf	—
12	$CQI < 7.5$	Node 7 (Y)	Leaf	Leaf	—
13	— (leaf)	Node 7 (N)	—	—	No_HO (n = 125: 5P/120NH)
14	— (leaf)	Node 8 (Y)	—	—	No_HO (n = 13: 1P/12NH)
15	— (leaf)	Node 8 (N)	—	—	No_HO (n = 27: 5P/22NH)
16	$SNR < -3.5$	Node 10 (Y)	Leaf	Leaf	—
17	$RSRQ\_delta < -0.5$	Node 10 (N)	Leaf	Node 25	—
18	— (leaf)	Node 11 (Y)	—	—	No_HO (n = 14: 0P/14NH)
19	— (leaf)	Node 11 (N)	—	—	No_HO (n = 29: 6P/23NH)
20	— (leaf)	Node 12 (Y)	—	—	No_HO (n = 15: 4P/11NH)
21	— (leaf)	Node 12 (N)	—	—	No_HO (n = 25: 1P/24NH)
22	— (leaf)	Node 16 (Y)	—	—	Pending_HO (n = 10: 8P/2NH)
23	— (leaf)	Node 16 (N)	—	—	No_HO (n = 24: 4P/20NH)
24	— (leaf)	Node 17 (Y)	—	—	Pending_HO (n = 62: 54P/8NH)
25	$RSRP\_delta < 2.5$	Node 17 (N)	Node 26	Leaf	—
26	$Speed < 11.5$	Node 25 (Y)	Node 28	Node 29	—
27	— (leaf)	Node 25 (N)	—	—	Pending_HO (n = 13: 13P/0NH)
28	$NRxRSRP < -101$	Node 26 (Y)	Node 30	Leaf	—
29	$RSRP\_std\_before < 1.66$	Node 26 (N)	Leaf	Node 33	—
30	$RSSI < -92.5$	Node 28 (Y)	Leaf	Leaf	—
31	— (leaf)	Node 28 (N)	—	—	Pending_HO (n = 19: 10P/9NH)
32	— (leaf)	Node 29 (Y)	—	—	No_HO (n = 12: 1P/11NH)
33	$RSRP\_std\_before < 3.28$	Node 29 (N)	Leaf	Leaf	—
34	— (leaf)	Node 30 (Y)	—	—	Pending_HO (n = 10: 7P/3NH)
35	— (leaf)	Node 30 (N)	—	—	Pending_HO (n = 14: 14P/0NH)
36	— (leaf)	Node 33 (Y)	—	—	Pending_HO (n = 17: 13P/4NH)
37	— (leaf)	Node 33 (N)	—	—	No_HO (n = 14: 6P/8NH)

To compare with baseline cases, the paper looks at the literature. Standard 3GPP A3 event-triggered handover mechanism relies only on RSRP comparisons with fixed Handover Margin (HOM) and Time-to-Trigger (TTT) parameters. Studies have shown that this approach performs poorly in dense heterogeneous environments. For instance, Shayea *et al.* (Shayea *et al.*, 2025) demonstrates that MLP-based handover optimization outperforms the A3 method by approximately 79.8% in ping-pong reduction and 76.6% in handover failure reduction in ultra-dense small-cell HetNets, confirming that fixed-threshold methods are inadequate for modern networks. In addition, de Brito Guerra *et al.* (de Brito Guerra, Dantas & Sousa Jr., 2023) shows that state-of-the-art RSRP-only baselines achieve an F1-score below 40%, compared to multi-feature ML models that can exceed 70% accuracy. The proposed Decision Tree model, which uses 17 features (13 engineered + a 4-feature NRx neighbor-cell block) and achieves an F1-score of 0.642 with an AUC of 0.797, significantly outperforms RSRP-only

approaches while maintaining interpretability and low computational cost compared to deep learning alternatives. Note: the F1/AUC values reported alongside the literature comparison above refer to different datasets and measurement protocols; see Table 5 for the head-to-head LOTO numbers on the present dataset. The same LOTO protocol is used to evaluate in-house baselines: a Decision Tree without the neighbor-cell (NRx) block achieves a F1 = 0.593 and AUC = 0.741, so adding the NRx block gives a +0.049 F1 lift. Among simple threshold rules, RSRP < -100 dBm yields F1 = 0.519,  $\Delta$ RSRP < -3 dB yields F1 = 0.377, and an A3-like NRx margin rule yields F1 = 0.590. The proposed Decision Tree with the NRx block reaches F1 = 0.642 and AUC = 0.797, giving a 0.052 F1 increase over the best threshold rule, while remaining causal and deployable in real time.

Some complementary baselines have been identified as future work such as (i) a logistic regression, (ii) a random-forest, (iii) a gradient-boosted tree, and (iv) a 3GPP A3 reproduction with specified HOM, TTT, and hysteresis values run on the same LOTO folds.

### 5.2.2. Ping-Pong Handling

Ping-pong events are identified using the 3GPP TR 36.839 dwell-time and oscillation rules and excluded from the positive class before training. Table 7 reports the labeling breakdown: of the 206 valid-window cell-change events, 167 are labeled as genuine handovers and 39 as ping-pong. Records on 27 January 2018 at 10:58:49 show the highest rate of ping-pong (41.2%), consistent with the UE traversing a cell-overlap region, whereas the trace from 30 November 2017 at 16:48:26 contains only a single ping-pong event, consistent with the UE remaining in a single cell with no overlap.

Table 7. Ping-pong labeling statistics.

Classification	Count	Percentage
Genuine handover (positive class)	167	81.07%
Ping-pong (excluded from training)	39	18.93%

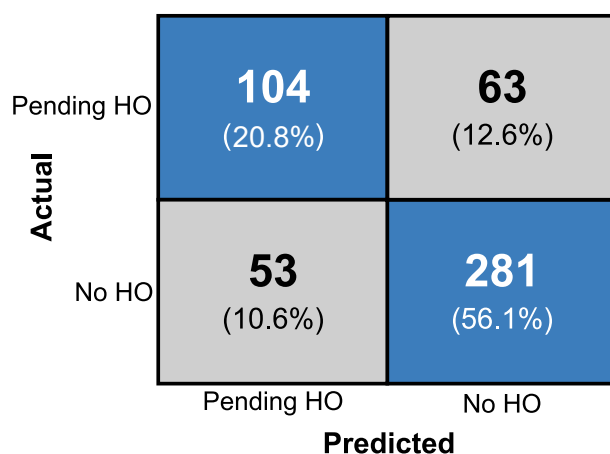


Figure 2. Confusion matrix of the proposed single-stage classifier under Leave-One-Trace-Out cross-validation. The 62.3% on the true-positive cell denotes recall ( $TP/(TP+FN) = 104/167$ ); the remaining percentages are fractions of the total 501 evaluated samples.

### 5.3. Visualizations

To showcase the findings, the proposed handover system generates figures that help interpret and visualize the results.

### 5.3.1. Decision Tree Structure

Figure 1 shows the top three levels of the trained decision tree, while Table 6 provides the complete node-by-node specification of all 18 split nodes and 19 leaf nodes. Each internal node is described with a split feature and threshold, and each leaf reports the assigned class together with the within-leaf training-sample breakdown. The displayed tree corresponds to a representative LOTO fold (481 training samples; one trace held out from the full 501-sample set). The root-level split on *NRxRSRP\_missing* partitions samples by neighbor-cell measurement availability: when neighbor information is present (Yes branch,  $n = 305$ ) the tree first inspects neighbor-cell quality ( $NRxRSRQ < -18.5$ ) and only then evaluates serving-signal and CQI features; when neighbor information is absent (No branch,  $n = 176$ ) the tree falls back to serving-cell quality ( $RSRQ < -14.5$ ). This structure operationalises the design intent of the NRx feature block: neighbor-cell awareness is used as the primary discriminator whenever it is available, while a serving-only path provides graceful degradation otherwise.

### 5.3.2. Feature Importance

Figure 3 shows the ranking of the 17 features by their importance to the model. The importance stems from their Gini impurity reduction across all splits, averaged over the 16 LOTO folds. *CQI* dominates the ranking, reflecting its role as a high-information physical-layer indicator that aggregates SINR, modulation, and coding margin into a single index that degrades rapidly when a handover is imminent. *NRxRSRQ* comes second, confirming that neighbor-cell quality is the strongest single predictor of an impending cell change once such information is available. *RSRP\_delta* ranks third, capturing the short-horizon serving-signal trend; the remaining contributors – *SNR*, *NRxRSRP\_margin*, *RSRQ*, *RSRP\_std\_before*, *RSRP*, and *NRxRSRP\_missing* – each add complementary radio-condition or neighbor-availability information. The mobility features *Speed* and *ServingCell\_Distance* carry small but non-zero weight, indicating that the tree uses them mainly inside deeper subtrees rather than near the root. Three features – *ServingCell\_Distance\_delta*, *NetworkMode*, and *dist\_delta\_missing* – contributed zero importance and are candidates for removal in a leaner deployed model; their high autocorrelation with already-included signals likely explains the redundancy.

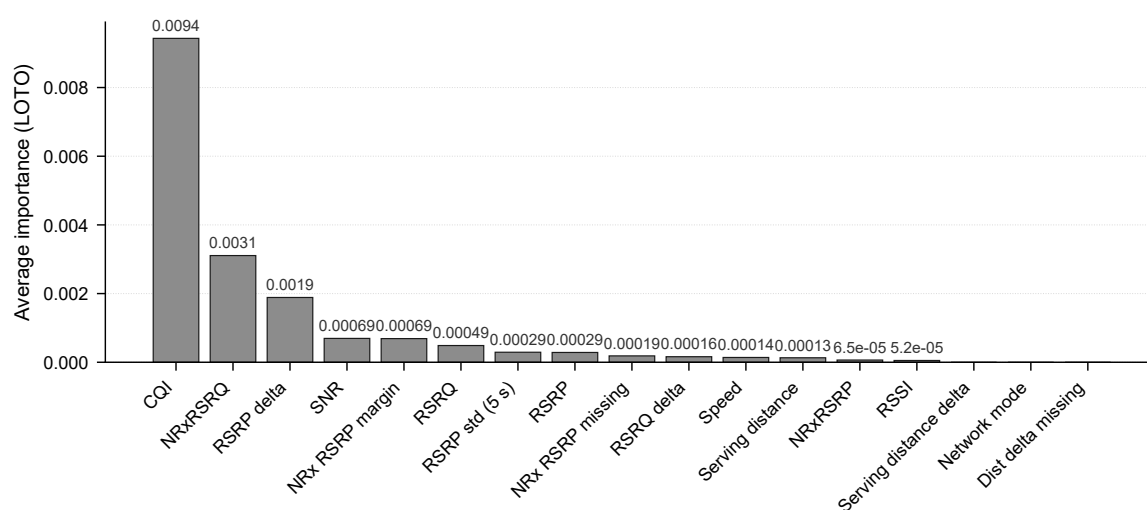


Figure 3. Feature importance ranked by Gini impurity reduction in the full-data Decision Tree.

### 5.3.3. Confusion Matrix and Pipeline Flow

Figure 2 shows the LOTO confusion matrix with TP = 104, FN = 63, FP = 53, and TN = 281. The pipeline is a single block that flows as follows: it starts with a pre-handover window, then a feature vector comprising 17 features (13 engineered + the 4-feature NRx neighbor-cell block), then the decision tree, producing a direct prediction of a genuine handover.

### 5.3.4. ROC Curve

Figure 4 shows the Receiver Operating Characteristic curve obtained from the accumulated LOTO prediction scores. The AUC of 0.797 indicates strong classifier discrimination. The curve's steep initial rise demonstrates that the model achieves high sensitivity at low false-positive rates, which is good for deployment where false alarms must be minimized.

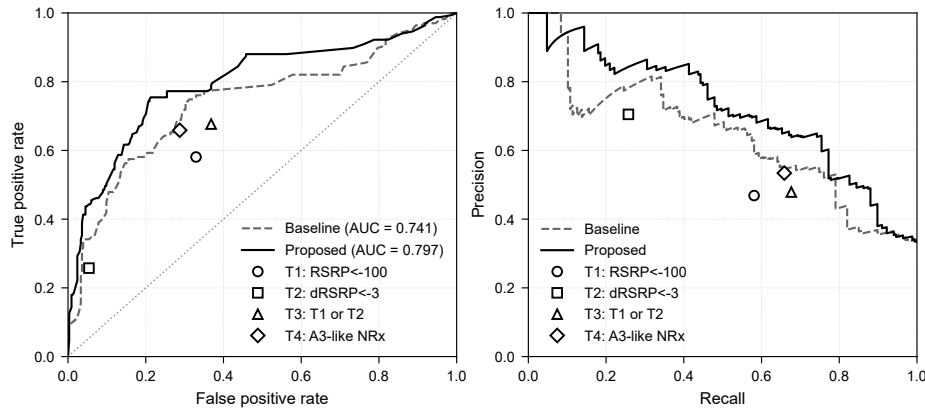


Figure 4. ROC (left) and Precision–Recall (right) curves from Leave-One-Trace-Out validation; AUC = 0.797.

## 5.4. KPI Analysis

### 5.4.1. Handover Probability - HOP

HOP is obtained as  $HOP = N_{HO,raw} / N_{total} = 211 / 10783 = 0.0196$ . The total probability of handover is 1.96%, which corresponds to 13.2 handovers per trace. Operator B has a higher HOP of 2.49% (97 HOs across 3,890 samples) compared to 1.65% for Operator A (114 HOs across 6,893 samples). This difference suggests that Operator B deploys smaller cells.

### 5.4.2. Ping-Pong Handover Probability - HOPP

HOPP is obtained as  $HOPP = N_{PP} / N_{HO,valid} = 39 / 206 = 0.1893$ , where  $N_{HO,valid} = 206$  counts cell-change events that have at least one measurement inside the 1–3 s prediction window. Of those 206 detected handovers, 39 (18.93%) are classified as ping-pong. All 39 are intra-RAT with no inter-RAT ping-pong. The effective HOPP is  $HOPP_{eff} = N_{PP,predicted} / N_{HO,actionable}$ , where  $N_{PP,predicted}$  counts cell-change events that are actually ping-pong but still predicted as genuine handovers.  $HOPP_{eff}$  is expected to be near zero since the system is designed to avoid ping-pong and unlearn them.

### 5.4.3. Radio Link Failure Rate - RLF

RLF is approximated by investigating where  $RSRP \leq -120$  dBm ( $Q_{out}$ ) persists for  $\geq 3$  consecutive seconds:

$$RLF_{rate} \approx N_{RLF} / N_{HO,raw} = 64 / 211 = 0.3033 \quad (15)$$

A total of 219 samples (2.03%) fall below the  $Q_{out}$  threshold, forming 64 sustained episodes; here  $N_{HO,raw} = 211$  refers to all detected cell-change events (including those without a valid prediction window), in contrast to  $N_{HO,valid} = 206$  used for HOPP. Also, 21 connected-to-idle (D→I) state transitions occur within 5 s of a handover, indicating connection drops. The mean RSRP at the handover instant is  $-99.8$  dBm; however, 18.0% of handovers occur at  $RSRP \leq -110$  dBm and 0.9% at  $RSRP \leq -120$  dBm, suggesting that a subset of handover decisions are triggered too late. The proposed system mitigates RLF risk by providing 1–3 s advance warning. Because the classifier is trained on a positive class that excludes ping-pong events, ping-pong transitions are avoided at the prediction stage. Each avoided ping-pong eliminates two handover transitions, so  $\Delta N_{RLF} = 2 \times N_{PP,avoided} \times P_{fail}$ . For an LTE per-handover failure probability  $P_{fail} = 2\%$ , avoiding all 39 ping-pongs amounts to an estimated  $\approx 1.5$  RLF events spared (order-of-magnitude; the figure scales linearly with the assumed

$P_{\text{fail}}$ ). The 2% value is an indicative LTE figure; handover-failure rates reported in the literature (Xenakis *et al.* (Xenakis, Passas, Merakos & Verikoukis, 2014)) are between 1–5% depending on UE speed, cell radius, and HOM/TTT configuration.

#### 5.4.4. Data Rate Impact

Throughput impact is identified by comparing the mean DL bitrate and UL bitrate across three conditions. The throughput drop is defined as  $\Delta\text{DR} = (\text{DR}_{\text{stable}} - \text{DR}_{\text{HO}}) / \text{DR}_{\text{stable}} \times 100\%$ . During stable periods (8,148 samples, >10 s from any HO), the mean DL throughput is 9,787.9 kbps. Near all handover events (1,644 samples,  $\pm 5$  s window), it remains 9,787.1 kbps, resulting in  $\Delta\text{DR}_{\text{all}} = 0.0\%$ . The remaining rows of the 10,783-row dataset fall in the 5–10 s buffer between the “stable” and “near-HO” regions and are excluded from this analysis to avoid double counting. Correct handovers cause virtually no throughput loss. However, isolating the 180 samples near ping-pong events reveals a severe penalty:  $\Delta\text{DR}_{\text{PP}} = 30.7\%$  for downlink and  $\Delta\text{UL}_{\text{PP}} = 31.4\%$  for uplink. Ping-pong handovers cause a 30.7% DL and 31.4% UL throughput degradation. This confirms that ping-pong events are the main source of throughput loss caused by mobility. Per-operator baseline data rates (computed over each operator’s own stable subset of the 8,148 stable samples) differ substantially: Operator B achieves 13,493 kbps versus 8,034 kbps for Operator A, likely reflecting differences in spectrum allocation or load.

**Table 8.** KPI Results Summary (dataset-level KPIs; the classifier reduces their occurrence at prediction time).

KPI	Overall	Op. A	Op. B
HOP	1.96%	1.65%	2.49%
HOPP	18.93%	—	—
RLF Rate	30.33%	—	—
$\text{DR}_{\text{stable}}$ (kbps)	9,788	8,034	13,493
$\Delta\text{DR}_{\text{all}}$	0.0%	—	—
$\Delta\text{DR}_{\text{PP}}$	30.7%	—	—

**Table 9.** Data Rate by Condition

Condition	Samples	DL (kbps)	UL (kbps)	DL Drop
Stable	8,148	9,788	179	—
Near all HOs	1,644	9,787	180	0.0%
Near PP HOs	180	6,780	123	30.7%

## 6. Limitations and Threats to Validity

Limitations present themselves strongly when interpreting the results. Regarding data era and radio technology, the extension to 5G and 6G is crucial since 5G is highly commercialized and actively researched, and further applicability work is needed; the elements that are transferable require more inspection. Regarding geographic scope, the traces confirm overlapping routes with two operators in one geographic region, and external validity across cities, topologies, and operators has not been assessed yet. The ping-pong rule uses information about the UE behavior in the 3-second window around a cell change; however, this cannot be replicated at inference time. Accuracy, computational efficiency, and deployment simplicity could be improved by exploring other machine learning models such as Random Forest, gradient-boosted trees, and logistic regression. The  $\text{RSRP}_{\text{std\_before}}$  feature is computed over a 5-sample backward window at 1 Hz, which introduces an inherent  $\sim 5$  s feature-assembly latency. Finally, the KPI data-rate and RLF impact estimates ( $\Delta\text{DR}_{\text{PP}}$ ,  $\Delta N_{\text{RLF}}$ ) are dataset-level observations combined with assumed failure probabilities.

## 7. Conclusion and Future Work

In this study, we used a Decision Tree model to support handover prediction using real network measurement data. The main objective of this work was to examine whether a lightweight machine learning model can help improve the handover decision process in dynamic cellular network environments. Our results show that the Decision Tree model can predict many handover events while also identifying many normal network situations. The confusion matrix shows that the model correctly predicted a number of handovers, although some false alarms and missed events still appear. This is expected because handover events are relatively rare compared to normal network states in our dataset, which makes the prediction task more difficult. We also observed that some features have a stronger influence on the decision process. In particular, the channel quality indicator (CQI), neighbor-cell quality (NRxRSRQ), and short-horizon serving-signal change (RSRP\_delta) ranked as the most informative features for predicting handover events, with additional contributions from SNR, the NRxRSRP margin, and signal-instability measures. This shows that using several network measurements together can help describe user mobility behavior better and make the handover decision more reliable. In future work, the model can be improved in different ways. One possible step is to extend the current Decision Tree model to a Random Forest model. The Random Forest algorithm uses several decision trees instead of only one tree, which may give more stable predictions and better performance in networks where conditions change quickly. Another way is to study how predictive handover models may work together with dual connectivity in modern cellular networks. Dual connectivity lets user devices connect to more than one base station at the same time. This can help reduce handover failures and keep the service more stable when users move between cells. These ways may help improve the reliability of handover prediction and make better mobility management in future cellular networks. The same idea could also be studied in future wireless networks such as the 6th generation, where handover management may become more challenging.

**Author Contributions:** Conceptualization, M.H., L.Y., and I.S.; methodology, M.H. and L.Y.; software, M.H. and L.Y.; validation, M.H., L.Y., and I.S.; formal analysis, M.H. and L.Y.; investigation, M.H. and L.Y.; data curation, M.H. and L.Y.; writing—original draft preparation, M.H. and L.Y.; writing—review and editing, L.Y., I.S., and L.R.; visualization, M.H. and L.Y.; supervision, I.S. and L.R.; project administration, I.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable. The drive-test measurements analyzed in this study did not involve human or animal subjects.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The drive-test traces analyzed in this study are available from the corresponding author upon reasonable request. The MATLAB implementation of the decision-tree pipeline, the LOTO evaluation harness, and the pre-handover-window labeling script will be released at a public repository upon acceptance (URL to be added).

**Acknowledgments:** The authors thank the anonymous reviewers for their constructive comments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Mohsin, H. S., Saad, W. K., & Shayea, I. (2023). Literature Review of Handover Decision Algorithms in 5G Networks. *2023 10th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, Istanbul, Turkiye, 1–6. <https://doi.org/10.1109/WINCOM59760.2023.10322978>
- Zaid, M., Kadir, M. K. A., Shayea, I., & Mansor, Z. (2024). Machine learning-based approaches for handover decision of cellular-connected drones in future networks: A comprehensive review. *Engineering Science and Technology, an International Journal*, 55, 101732. <https://doi.org/10.1016/j.jestch.2024.101732>

- Ahmed, A., Boulahia, L. M., & Gäiti, D. (2014). Enabling Vertical Handover Decisions in Heterogeneous Wireless Networks: A State-of-the-Art & A Classification. *IEEE Communications Surveys and Tutorials*, 16(2), 776–811. <https://doi.org/10.1109/SURV.2013.082713.00141>
- Loutfi, S. I., Shayea, I., Tureli, U., El-Saleh, A. A., Tashan, W., & Caglar, R. (2025). Machine learning for handover decision with mobile edge computing in 6G mobile network: a survey. *Engineering Science and Technology, an International Journal*, 69, 102131. <https://doi.org/10.1016/j.jestch.2025.102131>
- Ahmad, R., Sundararajan, E., Othman, N. E., & Ismail, M. (2017). Handover in LTE-advanced wireless networks: State of art and survey of decision algorithm. *Telecommunication Systems*, 66(3), 533–558.
- Amirova, A., Shayea, I., Yedilkhan, D., Aldasheva, L., & Zakirova, A. (2025). Handover Decisions for Ultra-Dense Networks in Smart Cities: A Survey. *Technologies*, 13(8), 313. <https://doi.org/10.3390/technologies13080313>
- Jahandar, S., Shayea, I., Gures, E., El-Saleh, A., Ergen, M., & Alnakhli, M. (2025). HOD with multi-access edge computing in 6G networks: A survey. *Results in Engineering*, 25, 103934.
- Elhilali, N., Badri, M., & Bouami, M. F. (2023). An overview of vertical handover decision algorithms. *AIP Conf. Proc.*, 2814(1), 030002. <https://doi.org/10.1063/5.0149277>
- Goh, M. I., Mbulwa, A. I., Yew, H. T., Kiring, A., Chung, S. K., Farzamnia, A., Chekima, A., & Haldar, M. K. (2023). Handover Decision-Making Algorithm for 5G Heterogeneous Networks. *Electronics*, 12, 2384. <https://doi.org/10.3390/electronics12112384>
- Xenakis, D., Passas, N. I., Merakos, L. F., & Verikoukis, C. V. (2014). Mobility Management for Femtocells in LTE-Advanced: Key Aspects and Survey of Handover Decision Algorithms. *IEEE Communications Surveys & Tutorials*, 16, 64–91.
- Gupta, A. K., Goel, V., Garg, R. R., Thirupurasundari, D. R., Verma, A., & Sain, M. (2021). A Fuzzy Based Handover Decision Scheme for Mobile Devices Using Predictive Model. *Electronics*, 10, 2016. <https://doi.org/10.3390/electronics10162016>
- Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and Regression Trees*. Wadsworth.
- 3GPP. (2013). Mobility Enhancements in Heterogeneous Networks (TR 36.839 Release 11).
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)* (Vol. 2, pp. 1137–1143). Montreal, Canada.
- Shayea, I., Ergen, M., Azmi, M. H., Colak, S. A., Nordin, R., & Daradkeh, Y. I. (2025). A Handover Decision Optimization Method Based on Data-Driven MLP in 5G Ultra-Dense Small Cell HetNets. *Journal of Network and Systems Management*, 33(2), 33. <https://doi.org/10.1007/s10922-025-09903-6>
- de Brito Guerra, T. C., Dantas, Y. R., & Sousa Jr., V. A. (2023). Deep Learning-Based Handover Prediction for 5G and Beyond Networks. In *Proc. IEEE Int. Conf. Commun. (ICC)* (pp. 1–6).
- Raca, D., Leahy, D., Sreenan, C. J., & Quinlan, J. J. (2018). Beyond Throughput: A 4G LTE Dataset with Channel and Context Metrics. In *Proceedings of the 9th ACM Multimedia Systems Conference (MMSys'18)*, Amsterdam, The Netherlands (pp. 460–465). <https://doi.org/10.1145/3204949.3208123>
- 3GPP. (2023). NR; Radio Resource Control (RRC); Protocol Specification. Technical Specification TS 38.331, Release 17, 3rd Generation Partnership Project.
- 3GPP. (2023). Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol Specification. Technical Specification TS 36.331, Release 17, 3rd Generation Partnership Project.
- Tayyab, M., Gelabert, X., & Jäntti, R. (2019). A Survey on Handover Management: From LTE to NR. *IEEE Access*, 7, 118907–118930. <https://doi.org/10.1109/ACCESS.2019.2937405>
- Polese, M., Giordani, M., Mezzavilla, M., Rangan, S., & Zorzi, M. (2017). Improved Handover Through Dual Connectivity in 5G mmWave Mobile Networks. *IEEE Journal on Selected Areas in Communications*, 35(9), 2069–2084. <https://doi.org/10.1109/JSAC.2017.2720338>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.