

## Article

# A Hybrid User Mobility Prediction Approach for Handover Management in Mobile Networks

Nasrin Bahra \*  and Samuel Pierre

Department of Computer and Software Engineering, Polytechnique Montreal, 2500, Chemin de Polytechnique, Montreal, QC H3T 1J4, Canada; samuel.pierre@polymtl.ca

\* Correspondence: nasrin.bahra@polymtl.ca

**Abstract:** Mobile networks are expected to face major problems such as low network capacity, high latency, and limited resources but are expected to provide seamless connectivity in the foreseeable future. It is crucial to deliver an adequate level of performance for network services and to ensure an acceptable quality of services for mobile users. Intelligent mobility management is a promising solution to deal with the aforementioned issues. In this context, modeling user mobility behaviour is of great importance in order to extract valuable information about user behaviours and to meet their demands. In this paper, we propose a hybrid user mobility prediction approach for handover management in mobile networks. First, we extract user mobility patterns using a mobility model based on statistical models and deep learning algorithms. We deploy a vector autoregression (VAR) model and a gated recurrent unit (GRU) to predict the future trajectory of a user. We then reduce the number of unnecessary handover signaling messages and optimize the handover procedure using the obtained prediction results. We deploy mobility data generated from real users to conduct our experiments. The simulation results show that the proposed VAR-GRU mobility model has the lowest prediction error in comparison with existing methods. Moreover, we investigate the handover processing and transmission costs for predictive and non-predictive scenarios. It is shown that the handover-related costs effectively decrease when we obtain a prediction in the network. For vertical handover, processing cost and transmission cost improve, respectively, by 57.14% and 28.01%.

**Keywords:** mobility prediction; machine learning; mobile networks



**Citation:** Bahra, N.; Pierre, S. A Hybrid User Mobility Prediction Approach for Handover Management in Mobile Networks. *Telecom* **2021**, *2*, 199–212. <https://doi.org/10.3390/telecom2020013>

Academic Editor: Alessandro Pozzebon

Received: 14 March 2021

Accepted: 20 April 2021

Published: 6 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Nowadays, there is a significant increase in the number of mobile users around the world. It is estimated that this number will dramatically increase up to 5.7 billions by 2023 based on the Cisco annual record [1]. Consequently, a considerable amount of mobility data is generated by these users while moving everyday. We can fully exploit the potential of these produced data in many applications in order to extract frequent patterns and to derive invaluable insights. This fact opens up plenty of golden opportunities in many prospective areas such as mobility-aware services. User mobility prediction techniques can be applied to these available data to obtain a deeper understanding of users' mobility behaviours and their demands by using data mining and learning algorithms.

Foreseeable future mobile communication networks are rightly expected to support high data rates and seamless connectivity for a vast number of devices. They are also expected to avoid lengthy delays when providing services to mobile users in the network. However, there are some inherent technical challenges in mobile networks that require special attention when providing a high quality of services (QoSs) for users [2]. User mobility can pose tremendous challenges and needs to be managed appropriately. Some of these problems are heterogeneity, more frequent handovers, increase in handover-related costs, and increase in call dropping probability [3]. Therefore, it is crucial to have proper mobility management in the network.

A self-organizing network (SON) is a potential long-term solution to cope with some of the mentioned underlying issues [4]. These networks are growing greatly in popularity due to the fact that they are able to dynamically adapt to changes and to learn from past experiences. A SON is composed of three main functions: self-configuration, self-optimization, and self-healing. In the context of intelligent mobility management, user mobility prediction is a subsection of self-optimization functions with the main objective of predicting the future location of the user using learning techniques. Mobility awareness can make potentially outstanding contributions to many research areas related to mobile network management including resource allocation, handover management, and location-based services and, in general, can increase the quality of services.

When a user moves between different cells, the procedure of reassigning user resources to a new base station is called handover (HO). Handover is triggered based on the received signal condition or network load balancing. In dense 5G cellular networks, user mobility results in more repeated handovers and service interruptions. This can cause high signaling overhead, especially for mobile users at high velocities. Hence, handover management has a leading role in cellular network performance in order to support seamless connectivity for users. Therefore, improving HO procedure performance can be highly beneficial in reducing link failures, latency, and costs in a cellular networks. Modeling user mobility can significantly help the network obtain prior knowledge about a user's future trajectory and handovers. Therefore, deploying predictive handover management, the network can prepare the required services for users in advance to reduce latency and costs [5]. Having predicted the future HO of the user, handover preparation steps can be performed beforehand, and when HO is needed, the process can start from the execution phase. This can reduce the number of signaling messages needed to be exchanged and consequently results in lower HO latency.

There are a large number of works related to user mobility prediction and mobility-aware services for cellular networks in the literature. Despite valuable efforts, a generic reliable mobility predictor with a low prediction error for short-term and long-term predictions with low complexity is still missing. We believe that the key factor for a successful mobility model is to properly discover the existing hidden correlations in a user's movement history. When the model captures the dependencies in the prior user movement trajectories, it can provide a better understanding of user mobility behaviour and can make a long-term prediction. In this paper, our main objective is to deploy user mobility prediction as an invaluable tool to optimize the handover signaling procedure. First, we propose a hybrid trajectory prediction method based on statistical and learning models. The proposed mobility model exploits the vector autoregression (VAR) model [6] and gated recurrent unit (GRU) [7] to further capture correlations in the user's past movement history as the input samples. Then, having predicted the future user trajectory, we optimize the handover signaling procedure in order to reduce the handover signaling costs. We observe that the predictor accuracy is of great important in obtaining an accurate and effective impact on handover signaling optimization. The major contributions of our work is as follows:

- Proposing a mobility prediction technique to model user mobility behaviour. The proposed VAR-GRU model predicts the future trajectory (i.e., path) of a user. The core concept is to fully analyze the existing dependencies in a user's past trajectories and to extract general patterns in the data.
- Investigating the impact of user mobility prediction on the conventional handover signaling procedure. Handover processing and transmission costs are evaluated to compare the predictive and non-predictive scenarios.
- Conducting experiments on the user mobility data generated from real users to provide an in-depth analysis of the effectiveness of the proposed approach.

The rest of this paper is organized as follows. Section 2 provides a short review of some of the recent related works in this field. Section 3 presents the proposed approach in two subsections: the proposed user mobility prediction technique and handover procedure

analysis. Section 4 provides an analysis regarding HO signaling cost. In Section 5, we summarize the results of our experiments. Finally, the conclusion of our work is provided in Section 6.

## 2. Related Works

There are a vast number of works in the field of user mobility prediction and mobility-aware services in cellular networks. Generally, user mobility prediction can be categorized in two main parts: (1) mobility models that predict user future path or trajectory (i.e., regression task) and (2) mobility models that predict user next location (i.e., classification task). Our work belongs to the first category and we essentially predict the future trajectory of the user. In the following, we provide a short review of recent works from both categories.

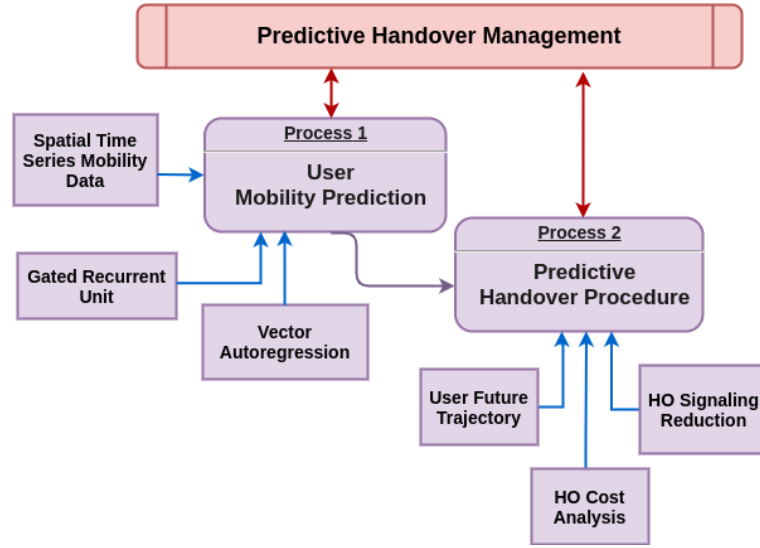
A huge part of the related works are based on the concept of machine learning algorithms. In [8], the authors introduced a hybrid mobility prediction method using principal component analysis (PCA) and gated recurrent unit. This approach was developed to predict Internet of Things (IoT) mobile users. A destination prediction approach was introduced in [9] using a long short-term memory (LSTM) network. This method investigates user prior knowledge in a bidirectional structure. In [10], a method based on deep learning techniques was adopted for multi-user trajectory prediction. This approach uses LSTM cells to learn a user's pattern and then extends it to the general case. In [11], an LSTM-based model was proposed for destination prediction. This model copes well with data sparsity as well. An RNN-based method was proposed in [12] for trajectory representation learning with the focus of considering both spatial and temporal features in trajectory learning. In [13], the authors proposed a mobility model based on the recurrent neural network variations. They proposed to deploy line simplification techniques to simplify the user trajectory. The core idea is to eliminate irrelevant data to reduce the execution time while improving the prediction accuracy. Then, the preprocessed dataset is used to learn the user mobility behaviour. In [14], the authors proposed a trajectory prediction method. Their technique explores existing relations between a user's morning and afternoon trajectories using a similarity metric. In [15], an analytical model was presented to calculate the handover-related costs including HO latency, signaling overhead, and call dropping. A part of the existing works is based on the popular concept of location-based social networks (LSBNs) that mainly investigates check-in datasets to predict the next point-of-interest (POI). In [16], the authors proposed a context-aware scheme to discover regular patterns in the user's movement history based on RNNs. To solve the data sparsity problem, they analyzed the social relationships as well. In [17], POI estimation was provided by applying gated recurrent units on a check-in dataset.

Moreover, there are many mobility models based on Markov models and their variations. In [18], the authors proposed a mobility prediction method based on Markov chains, with the main objective of reducing handover-related costs. In [19], the authors proposed a Markov-based model that predicts both a user's path and destination. In [20], a mobility model was introduced to predict the next base station in an LTE network using a Markov model. In [21], a mobility-aware proactive multi-cast technique was presented. This approach deployed a Markov model to estimate user's next cells and staying durations. However, it has been proven that Markov models fail to deal with radical changes in user mobility behaviour and to make a long-term prediction [22]. They cannot predict when hidden states increase. They performed more accurately when there were a limited number of observations.

We believe that a mobility model should be able to fully analyze the past trajectory of the user to properly predict the future path of that user with the smallest possible error. A model is needed that can perform effectively for both short- and long-term predictions with an acceptable error.

### 3. The Proposed Predictive Handover Management Approach

In this section, we present information regarding the proposed predictive handover management approach. Figure 1 shows the overall procedure of the handover management approach. This approach is composed of two main parts: (1) user mobility prediction and (2) predictive handover procedure analysis. For the first step, the main idea is to predict the future trajectory of the user using a hybrid model based on the VAR and GRU models. For the second step, we exploited the obtained prediction information from the last step to optimize the handover signaling procedure for the both horizontal and vertical handovers. In the following, we provide the details regarding each step.



**Figure 1.** The proposed handover management approach.

#### 3.1. User Mobility Prediction

*Objective:* Given the movement history of a user for the previous  $n$  time steps (i.e., user past trajectory), we want to predict the user's future trajectory for  $n$  time steps ahead.

We define the trajectory for user  $i$  as a sequence of time-stamped points,  $L_{useri} = (l_{t-n}, \dots, l_{t-2}, l_{t-1})$ , where  $l_t = (x_t, y_t, z_t)$  contains information on the geographical location of the user (i.e., longitude, latitude, and elevation).

First, we feed the user past trajectories into a vector autoregression model to obtain useful insights on the user mobility behaviour. Given  $L_t = (l_{t-n}, \dots, l_{t-2}, l_{t-1})$  as a user trajectory, in which  $L_t$  is a  $(v \times 1)$  vector with  $v$  variables, the VAR( $n$ ) model can be expressed as

$$l_t = a + C^1 l_{t-1} + C^2 l_{t-2} + \dots + C^n l_{t-n} + \epsilon_t \quad (1)$$

where  $C^1, C^2, \dots, C^n$  denote coefficient matrices that are square matrices of order  $v$  and  $a$  is an intercept vector.  $\epsilon_t$  is a white noise vector with zero mean, and it is assumed that there is no correlation between noises in the system across time. Each variable is obtained by the lagged values of all of the variables. In another words, to predict the value of a variable at the current time step, we consider the impact of all of the other variables from the past time steps. In our model, we considered three variables (i.e.,  $v = 3$ ) as defined earlier for the user trajectory  $l_t = (x_t, y_t, z_t)$ , and the model can be expressed as

$$\begin{pmatrix} x_t \\ y_t \\ z_t \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \begin{pmatrix} c_{11}^1 & c_{12}^1 & c_{13}^1 \\ c_{21}^1 & c_{22}^1 & c_{23}^1 \\ c_{31}^1 & c_{32}^1 & c_{33}^1 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{pmatrix} + \dots + \begin{pmatrix} c_{11}^n & c_{12}^n & c_{13}^n \\ c_{21}^n & c_{22}^n & c_{23}^n \\ c_{31}^n & c_{32}^n & c_{33}^n \end{pmatrix} \begin{pmatrix} x_{t-n} \\ y_{t-n} \\ z_{t-n} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{pmatrix} \quad (2)$$

Then, we have

$$\begin{aligned}x_t &= a_1 + c_{11}^1 x_{t-1} + c_{12}^1 y_{t-1} + c_{13}^1 z_{t-1} + \dots + c_{11}^n x_{t-n} + c_{12}^n y_{t-n} + c_{13}^n z_{t-n} + \epsilon_{1t} \\y_t &= a_2 + c_{21}^1 x_{t-1} + c_{22}^1 y_{t-1} + c_{23}^1 z_{t-1} + \dots + c_{21}^n x_{t-n} + c_{22}^n y_{t-n} + c_{23}^n z_{t-n} + \epsilon_{2t} \\z_t &= a_3 + c_{31}^1 x_{t-1} + c_{32}^1 y_{t-1} + c_{33}^1 z_{t-1} + \dots + c_{31}^n x_{t-n} + c_{32}^n y_{t-n} + c_{33}^n z_{t-n} + \epsilon_{3t}\end{aligned}\quad (3)$$

Since the VAR model is of order  $n$ , each variable is calculated by the  $n$  lagged values of all three variables in the model from the previous time steps. For example,  $c_{11}^1$  and  $c_{12}^1$  determine the impacts of  $x_{t-1}$  and  $y_{t-1}$ , respectively, when calculating  $x_t$ . This gives us a better understanding of input sample correlations.

For the next step, we feed the obtained results into a deep GRU-based neural network. For each time step, we consider the current location and the previous state to obtain the next state as the output. The GRU model has two main parts that deal with keeping or deleting user data from the past, namely the update gate and the reset gate. Using these two gates, the GRU model can solve the vanishing/exploding issue of RNNs and can make a long-term prediction [7]. Table 1 presents the model parameters.

The update gate  $u(t)$  can be obtained by

$$u(t) = \sigma(b_u + U_u L(t) + W_u h(t-1)) \quad (4)$$

In (4), user trajectory ( $L(t)$ ) and the information from the previous step ( $h(t-1)$ ) are multiplied by their weights (the weights control the impact of each value on the final decision). Then, the sum of this multiplication with the update bias value goes through a sigmoid function ( $\sigma$ ). The output is a value between 0 and 1. The update gate decides which part of the user's past movement history should be kept for analysis.

Next, the reset gate  $r(t)$  can be given by

$$r(t) = \sigma(b_r + U_r L(t) + W_r h(t)) \quad (5)$$

It has the same procedure as that in Equation (4). The only difference is that they are multiplied by their own weights. The reset gate decides which part of the past information should be eliminated for future decisions. Lastly, we can compute the effect of these gates in the output as follows:

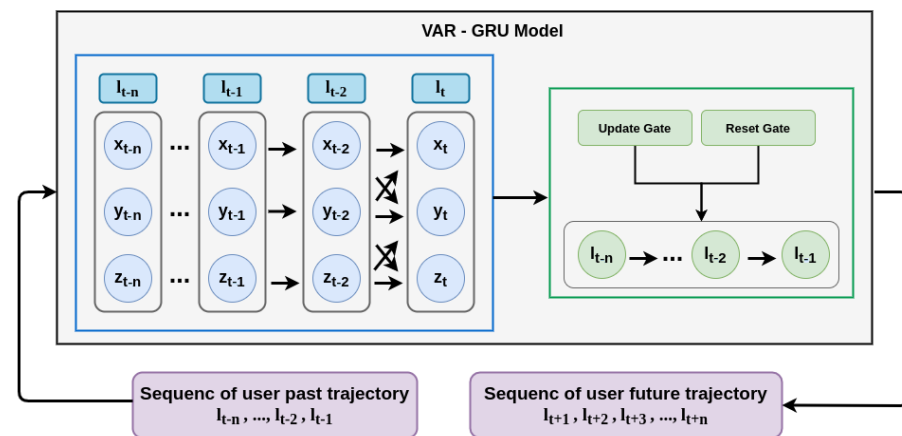
$$h(t) = u(t-1)h(t-1) + (1 - u(t-1))\sigma(b + U L(t-1) + W r(t-1)h(t-1)) \quad (6)$$

where  $b$ ,  $b_u$ , and  $b_r$  are bias vectors;  $U$ ,  $U_u$ , and  $U_r$  are input weight matrices; and  $W$ ,  $W_u$ , and  $W_r$  are recurrent weight matrices. Additionally,  $\sigma$  denotes a sigmoid function that has a value between 0 and 1 in the output.

**Table 1.** GRU model notations.

|               |                           |               |                       |
|---------------|---------------------------|---------------|-----------------------|
| $b, b_u, b_r$ | Bias vectors              | $U, U_u, U_r$ | Input weight matrices |
| $W, W_u, W_r$ | Recurrent weight matrices | $h(t)$        | Hidden states         |
| $u(t)$        | Update gate               | $r(t)$        | Reset gate            |

Figure 2 depicts the VAR-GRU trajectory prediction model. As shown, we took a sequence of the user past trajectory as the input. First, the VAR unit models the dynamic behaviour of the variables based on the raw data. Then, the obtained results are fed into the GRU network for the training and inference steps and produce the future trajectory of the user in the output.



**Figure 2.** The proposed VAR-GRU mobility model.

### 3.2. Predictive Handover Procedure

*Objective:* Given the user future trajectory, we want to reduce the number HO signaling messages needed when handover is triggered.

In this section, we investigate the conventional (i.e., non-predictive) and predictive handover signaling procedures. More specifically, we analyze the impact of trajectory prediction on handover signaling flow and compare it with the non-predictive handover scenario. A conventional handover procedure has three main steps including handover preparation, execution, and completion [23]. In the preparation step, target gNB is selected by the source gNB. The target gNB completes admission control. In the execution step, the user device attaches to the target gNB. In the last step, HO completion, the data path is switched to the new gNB and the handover steps are completed.

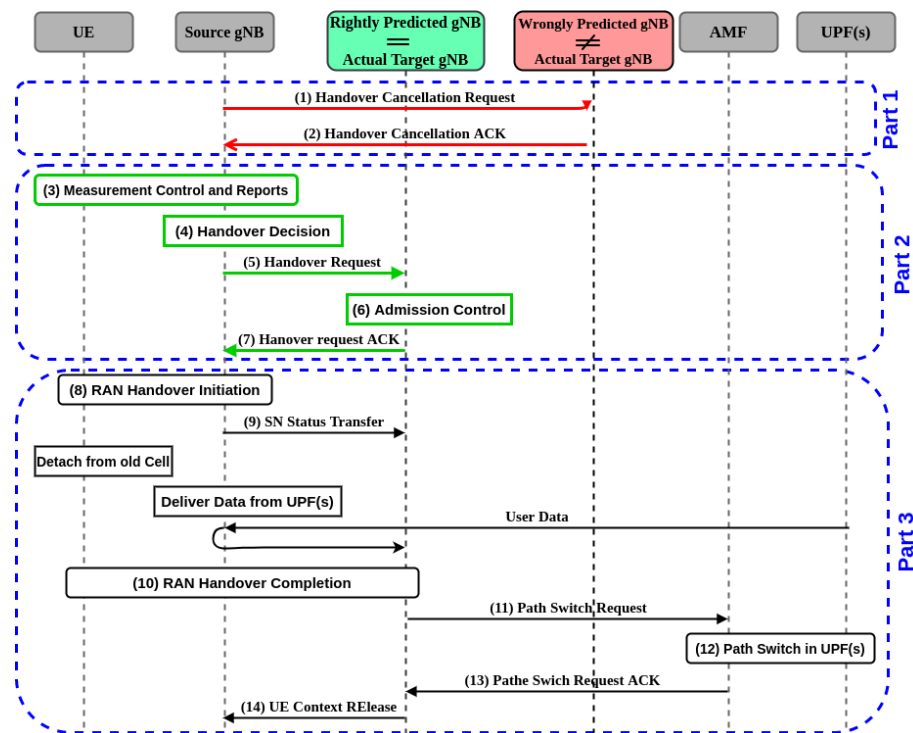
However, if the network has reliable information on the user's future locations (i.e., path), it can effectively perform some of the HO procedure steps before handover triggers. With correct prediction of the user future trajectory, we can complete the HO preparation phase in advance and when user requires a handover, it will start the execution phase. However, with a wrong prediction, even more signaling messages are required compared to the conventional HO. In the following, we elaborate on the impact of prediction on HO signaling in both horizontal and vertical handovers in a 5G architecture.

Figure 3 shows a 5G intra-AMF-UPF handover signaling process where we predicted the target base station. AMF and UPF in 5G architecture, respectively, refer to access and mobility management function (AMF) and user plane function (UPF). We considered both the cases of right and wrong predictions regarding a target gNB. As we can see in the figure, when the target gNB is correctly predicted, it is equal to the actual next base station (highlighted in green). However, with a wrong prediction, the predicted target base station is not the same as the actual next base station (highlighted in red). In this case, the source base station should send a message to the wrongly predicted base station and cancel the HO request. Therefore, there are three cases including (1) non-predictive HO (i.e., conventional HO), (2) predictive HO with right prediction, and (3) predictive HO with wrong prediction. For the first case, we need to perform all of the HO preparation, execution, and completion steps. According to Figure 3, signaling messages from numbers 3 to 14 need to be exchanged (i.e., parts 2 and 3). For the second case, when we predict the next base station correctly, the preparation steps can be performed in advance. When HO is triggered, the procedure starts with the execution phase. Thus, we need signaling messages from numbers 8 to 14 (i.e., part 3). Lastly, for the third case with a wrong prediction, signaling messages from numbers 1 to 14 are needed to complete the HO. First, resources at the wrongly predicted gNB should be released (part 1). Next, the non-predictive HO is applied to communicate with the actual next base station (parts 2 and 3).

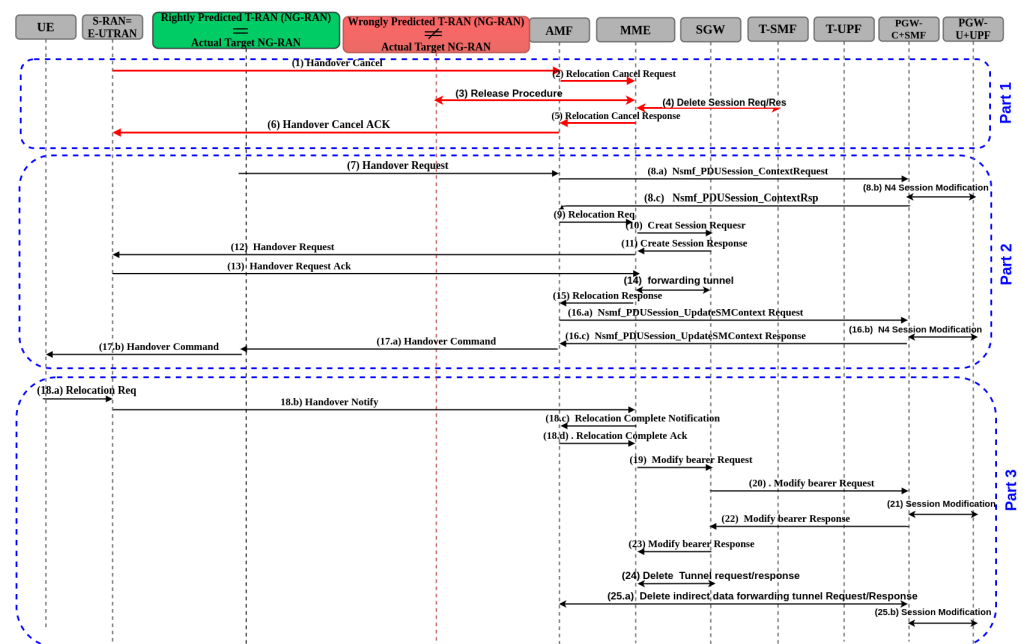
Similarly, Figure 4 shows an inter-radio access (RAT) handover signaling diagram from 5G next generation core (5G NGC) to evolved packet system (EPS). Similar to the



procedure in the Figure 3, we investigated the HO signaling procedure for right and wrong predictions regarding the next target gNB. There are three cases here as well. We note that the prediction accuracy has a vital role in the success of reducing HO signaling costs. An accurate prediction can reduce the signaling noticeably. On the other hand, as is evident, wrong predictions lead to even more signaling messages. Therefore, it is important to give undivided attention to the mobility predictor accuracy performance.



**Figure 3.** Intra-AMF-UPF handover signaling diagram. AMF and UPF in the 5G architecture refer to access and mobility management function (AMF) and user plane function (UPF).



**Figure 4.** Inter-RAT handover signaling diagram (from 5G NGC to EPS).

#### 4. Handover Signaling Cost Analysis

To investigate the impact of the proposed predictive handover procedure, we studied handover cost signaling for both the predictive and non-predictive (i.e., conventional handover procedure) cases. We assessed the handover signaling cost based on processing and transmission costs similar to that in [18,24,25].

##### 4.1. Processing Cost

We define the handover processing cost  $C_p$  based on the required number of signaling messages during a handover. Therefore, handover processing cost reduction can be given by

$$C_{P-Reduction} = \left[ \frac{1}{NM_{np}} (NM_{np} - NM_p) \right] \times 100\% \quad (7)$$

where  $NM_{np}$  and  $NM_p$  denote the number of messages needed for non-predictive and predictive scenarios, respectively.

##### 4.2. Transmission Cost

Transmitting handover messages causes an inevitable delay in establishing a link. The total HO messages  $M$  needed to perform a handover is as follows:

$$M = m_{prep} + m_{exe} + m_{comp} \quad (8)$$

where  $m_{prep}$ ,  $m_{exe}$ , and  $m_{comp}$  denote the messages that are needed to be exchanged during the HO preparation, execution, and completion steps, respectively. HO transmission cost  $C_T$  can be calculated based on the required link delay for handover signaling message transmission. This cost can be obtained by

$$C_T = \frac{1}{1ms} [D_M] \quad (9)$$

where  $D_M$  is the total required link delay for HO steps and can be defined as

$$D_M = \sum d_s, \quad s \in \mathbb{N} \quad (10)$$

where  $s$  refers to the number of signal (see Figures 3 and 4) and  $d_s$  denotes the link delay between two nodes regarding signal message  $s$ . In other words, HO transmission cost is calculated based on the total delay needed to establish links between nodes and to complete the handover process.

Thus, for an intra AMF-UPF handover, the transmission cost for non-predictive or conventional HO can be expressed as

$$C_{T-np} = \frac{1}{1ms} [D_{M-np}], \quad D_{M-np} = \sum_{s=3}^{14} d_s \quad (11)$$

where we calculate the delay for sending signal message 3 to 14 (according to Figure 3).

For the predictive HO, it is crucial to consider the impact of not only the right prediction but also the wrong prediction. Thus, we have the following:

$$C_{T-p} = \frac{1}{1ms} [D_{M-p}], \quad D_{M-p} = P_{acc}(D_{M-right}) + (1 - P_{acc})(D_{M-wrong}) \quad (12)$$

where  $P_{acc}$  is the accuracy of the prediction and  $D_{M-right}$  and  $D_{M-wrong}$  are, respectively, the link delay regarding signaling messages when we predicted the target base station correctly or incorrectly. They can be calculated as

$$D_{M-right} = \sum_{s=8}^{14} d_s, \quad D_{M-wrong} = \sum_{s=1}^{14} d_s \quad (13)$$



where we calculated the delay for sending signal message 8 to 14 for right predictions and 1 to 14 for wrong predictions (according to Figure 3). Here, we considered a horizontal handover (Figure 3) for the range of  $s$  in Equations (11) and (13). For vertical HO, we used the same procedure for cost calculation, and the range of  $s$  is defined according to Figure 4.

## 5. Experimental Results

In this section, we carried out a series of simulations to investigate the impact of the proposed approach. Our main objective was to evaluate the performance of the proposed mobility model and to then analyze the impact of the prediction on HO-related costs. In the following, first, we describe the mobility data that we used for our experiments and the approaches with which we compared our method. Next, we provide the results regarding the mobility model performance and its effect on HO procedure optimization.

### 5.1. Mobility Data Description

There are several human mobility data types including mobility data generated in cellular networks, WiFi networks, social networks, and global positioning system (GPS). We deployed GPS data since it has the highest localization accuracy. To conduct our experiments and to analyze the impact of the proposed method on different users, we chose mobility data from open street map (OSM). We deployed GPS traces that were uploaded by different users and are publicly available (<https://www.openstreetmap.org/traces>, accessed on 12 October 2020). For each user  $i$ , a sequence of past GPS trajectories  $L_{useri} = (l_{t-n}, \dots, l_{t-2}, l_{t-1})$  was provided as the input of the mobility model. A mobility predictor uses this information as prior knowledge of user mobility behaviour and tries to learn the repetitive patterns within. These sequential data are time-stamped data that contain information regarding time, longitude, latitude, and elevation of each location point in the user trajectory. Hence,  $l_t$  in a user input sequence that contains information on longitude ( $x_t$ ), latitude ( $y_t$ ), and elevation ( $z_t$ ) at time step  $t$  (i.e.,  $L_{useri} = ((x_{t_1}, y_{t_1}, z_{t_1}), (x_{t_2}, y_{t_2}, z_{t_2}), \dots)$ ).

### 5.2. Comparison

In order to provide a better understanding of how well our method works, we provide a fair comparison between some related techniques:

- Recurrent neural network [26]: a RNN-based mobility model analyzes the user's past locations sequentially.
- Long short-term memory [27]: LSTM-based approaches deploy state units and a forget gate to learn the mobility pattern.
- Gated recurrent unit [28]: GRU-based techniques control the impact of the latest observations using update and reset gates.

### 5.3. Experimental Settings

We considered the same hardware and software configurations for our method and all of the other methods. All of the simulations were performed using an Intel core i7-6700k CPU with 4.00 GHz and 32 GB RAM. We used keras for the simulations, which is a popular Python library. In the neural network, we deployed four layers with 100 neurons in each layer. To avoid overfitting, we used two drop out layers with values of 0.2 that randomly dropped 20 percent of the connections in each layer. Stochastic gradient decent (SGD) was used as the model optimizer with a learning rate of 0.001. Table 2 shows the simulation parameters that we considered for our experiments.

**Table 2.** Simulation parameters for training the proposed mobility model.

|                          |                |                 |           |
|--------------------------|----------------|-----------------|-----------|
| Number of layers         | 4              | Model optimizer | SGD       |
| Number of neurons        | 100            | Learning rate   | 0.001     |
| Weight initializer       | Glorot uniform | Loss function   | RMSE, MAE |
| Training data percentage | 75%            | Sequence length | 5–2000    |
| Batch size               | 10             | Dropout rate    | 0.2       |

#### 5.4. Mobility Model Performance

To evaluate the error of the mobility model, we considered two metrics: root mean square error (RMSE) and mean absolute error (MAE). Given  $b$  input samples from  $n$  total number of samples, RMSE can be given by

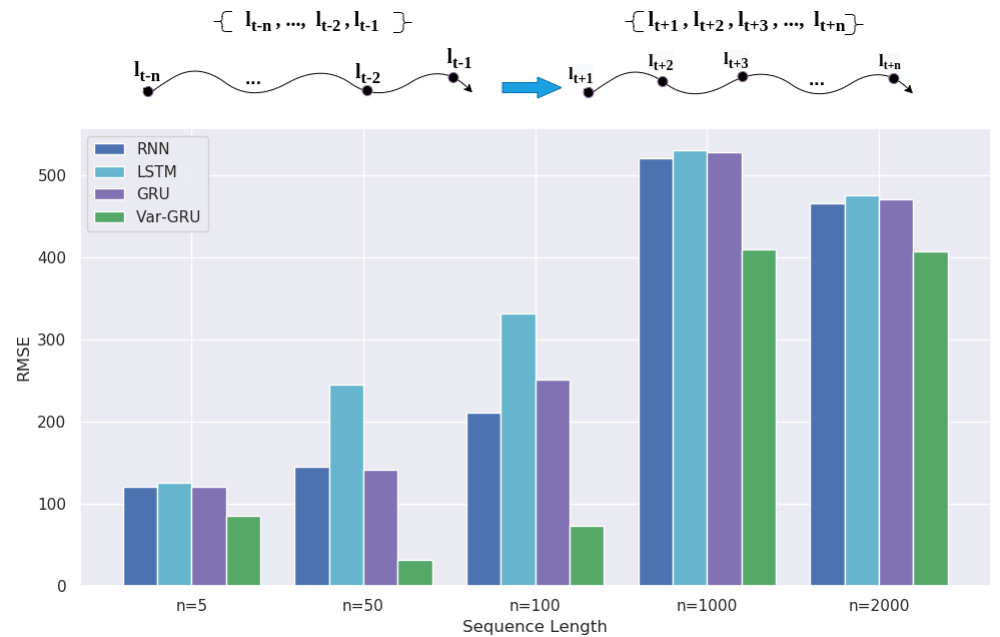
$$RMSE_b(\bar{y}) = \sqrt{1/n \sum_{i=1}^b (y(x) - \bar{y}(x))^2} \quad (14)$$

where  $x$ ,  $y(x)$ , and  $\bar{y}(x)$  are the input sample, actual result, and predicted result. The ideal value for RMSE is zero to show that the predicted and actual values are the same. Additionally, the MAE value can be calculated as

$$MAE_b(\bar{y}) = 1/n \sum_{i=1}^b |y(x) - \bar{y}(x)| \quad (15)$$

with lower values of MAE and RMSE being better for the predictor.

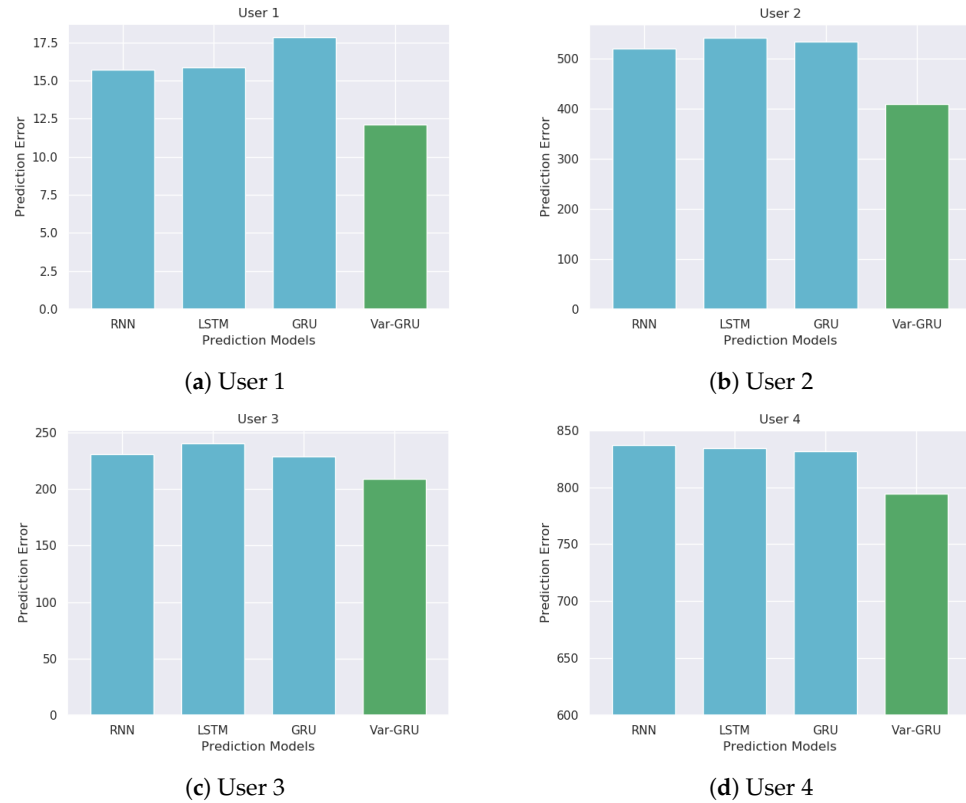
Length of the user sequence plays a vital role in evaluating a mobility model performance. Generally, when the length of the sequence is small, it is easier for the model to investigate the correlations. There is a tradeoff between sequence length and mobility behaviour learning. It is hard for the model to extract the user's mobility pattern when it deals with either a too short trajectory or a too long trajectory. Figure 5 shows the impact of different sequence lengths ( $n$ ) on the mobility model prediction error. Our model has a same-length input and same-length output structure. As shown, the proposed VAR-GRU mobility model has the lowest prediction error for all values of  $n$ .



**Figure 5.** Impact of the sequence length on the prediction error.

One important point for robust and reliable mobility models is the fact that users have different degrees of predictability based on their movement regularities [29]. Some users have more repetitive behaviour and thus are more predictable. Some users tend to have constantly changing mobility behaviour, and therefore, they are more unpredictable. A reliable mobility predictor should be able to model different users' mobility with low error. Figure 6 shows how the proposed mobility model works when dealing with four different users. As shown, the VAR-GRU model outperforms other methods for the four different users.

Next, we make a comparison between the mobility models based on the MAE value. Table 3 summarizes the results regarding each method. We observe that the proposed VAR-GRU trajectory predictor model outperforms the other approaches in terms of MAE value.



**Figure 6.** Impact of different mobility behaviours on the performance of the mobility models.

Next, having predicted the future trajectory of the user, we investigated the impact of this prediction on handover costs in terms of number of messages required and the transmission costs. After obtaining the predicted user future trajectory, we evaluated the accuracy of our predictions based on the actual future trajectory. We evenly distributed the base stations (BS) in the area of the user trajectory within the cell range of 1000 m. Therefore, each location point was within a cell area and we assigned it to the base station of that cell. We performed these steps for both the predicted and the actual user path. Then, we measured the accuracy of our predictions as a classification task and to observe whether the BS associated with the predicted trajectory points matched those of the actual trajectory points.

### 5.5. Impact of Prediction on HO Costs

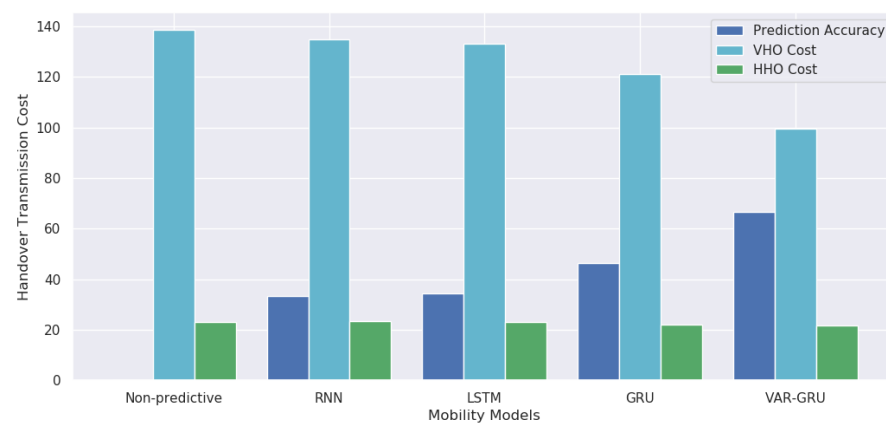
Figure 7 represents the results of the vertical and horizontal costs regarding the non-predictive case and the predictive case using different mobility models. We can observe the predictor accuracy and the corresponding HO costs for each model. Table 4 summarizes the delays regarding each type of link between two nodes in the process of HO completion. As shown, we obtained the highest transmission cost at 138.5 for the vertical handover when there is no prediction, and it gradually decreased as the prediction accuracy increased. It has the lowest value for the VAR-GRU at 99.7 (i.e., 28.01% reduction). For horizontal HO, the transmission cost for the non-predictive case is 23. This value slightly increased when we used the RNN and LSTM models as the mobility model at, respectively, 23.3 and 23.2. However, for GRU and the proposed VAR-GRU models, this cost was reduced to 22 and 21.67.

**Table 3.** Mean absolute error value (MAE) for all of the techniques.

| Users  | RNN    | LSTM   | GRU    | VAR-GRU |
|--------|--------|--------|--------|---------|
| User 1 | 12.55  | 11.75  | 12.10  | 5.35    |
| User 2 | 308.46 | 317.09 | 306.99 | 240.11  |
| User 3 | 121.74 | 124.01 | 117.41 | 121.50  |
| User 4 | 497.07 | 502.05 | 497.01 | 474.87  |

**Table 4.** The delay regarding each link type in rgw HO procedure [10,25].

| Link Type                  | Delay  |
|----------------------------|--------|
| UE to NG-RAN               | 1 ms   |
| NG-RAN to AMF              | 7.5 ms |
| AMF to SeMMu (PGW-C + SMF) | 1 ms   |
| SeMMu to S-GW              | 7.5 ms |
| SeMMu to PGW-U+UPF         | 7.5 ms |
| SeMMu to PCRF+PCF          | 7.5 ms |
| AMF to AMF                 | 15 ms  |
| SeMMu to PGW               | 7.5 ms |
| SeMMu to E-UTRAN           | 7.5 ms |
| E-UTRAN to UE              | 1 ms   |
| PGW to PCRF                | 7.5 ms |
| S-GW to PGW                | 7.5 ms |
| SeMMu to SGSN              | 1 ms   |
| SGSN to RNC                | 6 ms   |
| SGSN to S-GW               | 7.5 ms |
| SeMMu to SeMMu             | 15 ms  |

**Figure 7.** Handover transmission cost.

In Table 5, the handover processing costs for non-predictive (i.e., conventional) and predictive (proposed method) scenarios are presented. As mentioned earlier, this cost is defined based on the number of handover messages required for HO procedure completion. According to Table 5, for horizontal handover (see Figure 3), the proposed method effectively reduces the number of messages from 12 to 7 (i.e., 41.66% improvement). However, if the target gNB was predicted incorrectly, 14 messages are needed to complete the HO. Additionally, for a vertical handover (see Figure 4), the proposed method improves the processing cost by 57.14% (from 28 to 12) and it needs 34 messages for the case with a wrong prediction. Therefore, it is vital to consider the impact of wrong predictions in our evaluations as well. The obtained processing costs suggest that not having any prediction is more efficient than making decisions based on wrong predictions. Hence, predictor accuracy is of primary importance.

**Table 5.** Handover Processing cost for non-predictive and predictive scenarios when we use horizontal HO (HHO) and vertical HO (VHO).

| Handover Type | Approach  | Processing Cost |
|---------------|---|-----------------|
| HHO           | Conventional Handover Signaling                   | 12 Messages     |
|               | Proposed Handover Signaling with right prediction | 7 Messages      |
|               | Proposed Handover Signaling with wrong prediction | 14 Messages     |
| VHO           | Conventional Handover Signaling                   | 28 Messages     |
|               | Proposed Handover Signaling with right prediction | 12 Messages     |
|               | Proposed Handover Signaling with wrong prediction | 34 Messages     |

## 6. Conclusions

In this paper, we highlighted the significance of self-organizing networks for future network management using learning techniques. Our main objective was to deploy mobility prediction as a promising tool for mobile network management. We introduced a hybrid VAR-GRU mobility model to predict user future trajectory. The proposed mobility model was able to extract user mobility behaviour and repetitive patterns in their movement history. After predicting users' future trajectories, we effectively reduced the required number of handover signalings and optimized the HO signaling procedure based on our predictions.

Moreover, we investigated the handover signaling costs for the predictive and non-predictive scenarios to analyze the impact of mobility awareness on handover-related costs. The simulation results showed that the HO processing cost and transmission cost are reduced when we use mobility prediction. Our experiments indicate that the accuracy of prediction is of great importance and has a leading role in the effectiveness of the proposed method. It was shown that the proposed mobility model had the lowest error in comparison with the baseline methods.

For future work, the proposed hybrid VAR-GRU model can be exploited in many potential areas to further optimize cellular network services. Mobility-aware resource allocation, location-based services, and mobility-aware call admission control mechanisms are some of the promising applications for the proposed mobility prediction model.

**Author Contributions:** Conceptualization, methodology and analysis, N.B. and S.P.; writing—original draft preparation, N.B.; writing—review and editing, S.P.; supervision, S.P. 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.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. This data is available under the Open Database License at <https://www.openstreetmap.org/copyright> (accessed on 12 October 2020).

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

## References

1. Cisco. Cisco Annual Internet Report (2018–2023) White Paper. 2020. Available online: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html> (accessed on 9 March 2020).
2. Fazio, P.; De Rango, F.; Tropea, M. Prediction and QoS Enhancement in New Generation Cellular Networks With Mobile Hosts: A Survey on Different Protocols and Conventional/Unconventional Approaches. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 1822–1841. [CrossRef]
3. O'Connell, E.; Moore, D.; Newe, T. Challenges Associated with Implementing 5G in Manufacturing. *Telecom* **2020**, *1*, 48–67. [CrossRef]
4. Klaine, P.V.; Imran, M.A.; Onireti, O.; Souza, R.D. A Survey of Machine Learning Techniques Applied to Self-Organizing Cellular Networks. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 2392–2431. [CrossRef]

5. Di Taranto, R.; Muppirisetty, S.; Raulefs, R.; Slock, D.; Svensson, T.; Wymeersch, H. Location-Aware Communications for 5G Networks: How location information can improve scalability, latency, and robustness of 5G. *IEEE Signal Process. Mag.* **2014**, *31*, 102–112. [\[CrossRef\]](#)
6. Zivot, E.; Wang, J. Vector Autoregressive Models for Multivariate Time Series. In *Modeling Financial Time Series with S-Plus®*; Springer: New York, NY, USA, 2003; pp. 369–413. [\[CrossRef\]](#)
7. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; The MIT Press: Cambridge, MA, USA, 2016.
8. Adege, A.B.; Lin, H.; Wang, L. Mobility Predictions for IoT Devices Using Gated Recurrent Unit Network. *IEEE Internet Things J.* **2020**, *7*, 505–517. [\[CrossRef\]](#)
9. Lu, Z.; Zhao, F.; Luo, H.; Wang, Q.; Shao, W. A Destination Prediction Algorithm Using Spatial Temporal Bidirectional LSTM Networks; In Proceedings of the IPIN 2019, Pisa, Italy, 30 September–3 October 2019.
10. Wang, C.; Ma, L.; Li, R.; Durrani, T.S.; Zhang, H. Exploring Trajectory Prediction Through Machine Learning Methods. *IEEE Access* **2019**, *7*, 101441–101452. [\[CrossRef\]](#)
11. Xu, J.; Zhao, J.; Zhou, R.; Liu, C.; Zhao, P.; Zhao, L. Predicting Destinations by a Deep Learning based Approach. *IEEE Trans. Knowl. Data Eng.* **2021**, *33*, 651–666. [\[CrossRef\]](#)
12. Fu, T.Y.; Lee, W.C. Trembr: Exploring Road Networks for Trajectory Representation Learning. *ACM Trans. Intell. Syst. Technol. (TIST)* **2020**, *11*. [\[CrossRef\]](#)
13. Bahra, N.; Pierre, S. RNN-Based User Trajectory Prediction Using a Preprocessed Dataset. In Proceedings of the 2020 16th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Thessaloniki, Greece, 12–14 October 2020; pp. 1–6. [\[CrossRef\]](#)
14. Sadri, A.; Salim, F.; Ren, Y.; Shao, W.; Krumm, J.; Mascolo, C. What Will You Do for the Rest of the Day?: An Approach to Continuous Trajectory Prediction. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *2*, 1–26. [\[CrossRef\]](#)
15. Ozturk, M.; Gogate, M.; Onireti, O.; Adeel, A.; Hussain, A.; Imran, M.A. A novel deep learning driven, low-cost mobility prediction approach for 5G cellular networks: The case of the Control/Data Separation Architecture (CDSA). *Neurocomputing* **2019**, *358*, 479–489. [\[CrossRef\]](#)
16. Chen, Y.; Long, C.; Cong, G.; Li, C. Context-Aware Deep Model for Joint Mobility and Time Prediction. In Proceedings of the 13th International Conference on Web Search and Data Mining, WSDM '20, Houston, TX, USA, 3–7 February 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 106–114. [\[CrossRef\]](#)
17. Gao, Q.; Zhou, F.; Trajcevski, G.; Zhang, K.; Zhong, T.; Zhang, F. Predicting Human Mobility via Variational Attention. In Proceedings of the World Wide Web Conference, WWW '19, San Francisco, CA, USA, 13–17 May 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 2750–2756. [\[CrossRef\]](#)
18. Mohamed, A.; Onireti, O.; Hoseinitabatabaei, S.A.; Imran, M.; Imran, A.; Tafazolli, R. Mobility prediction for handover management in cellular networks with control/data separation. In Proceedings of the 2015 IEEE International Conference on Communications (ICC), London, UK, 8–12 June 2015; pp. 3939–3944. [\[CrossRef\]](#)
19. Nadembega, A.; Hafid, A.; Taleb, T. A Destination and Mobility Path Prediction Scheme for Mobile Networks. *IEEE Trans. Veh. Technol.* **2015**, *64*, 2577–2590. [\[CrossRef\]](#)
20. Farooq, H.; Imran, A. Spatiotemporal Mobility Prediction in Proactive Self-Organizing Cellular Networks. *IEEE Commun. Lett.* **2017**, *21*, 370–373. [\[CrossRef\]](#)
21. Vasilakos, X.; Al-Khalidi, M.; Siris, V.A.; Reed, M.J.; Thomos, N.; Polyzos, G.C. Mobility-based Proactive Multicast for Seamless Mobility Support in Cellular Network Environments. In Proceedings of the Workshop on Mobile Edge Communications, MECOMM '17, Los Angeles, CA, USA, 21–25 August 2017; ACM: New York, NY, USA, 2017; pp. 25–30. [\[CrossRef\]](#)
22. Kulkarni, V.; Mahalunkar, A.; Garbinato, B.; Kelleher, J.D. On the Inability of Markov Models to Capture Criticality in Human Mobility. In Proceedings of the Artificial Neural Networks and Machine Learning–ICANN 2019: Image Processing, Munich, Germany, 17–19 September 2019; Springer: Cham, Switzerland, 2019.
23. 5G. Procedures for the 5G System (5GS). (3GPP TS 23.502 version 15.4.1 Release 15) 2019, 1. Available online: [https://www.etsi.org/deliver/etsi\\_ts/123500\\_123599/123502/15.04.01\\_60/ts\\_123502v150401p.pdf](https://www.etsi.org/deliver/etsi_ts/123500_123599/123502/15.04.01_60/ts_123502v150401p.pdf) (accessed on 1 March 2019).
24. Jain, A.; Lopez-Aguilera, E.; Demirkol, I. Evolutionary 4G/5G Network Architecture Assisted Efficient Handover Signaling. *IEEE Access* **2019**, *7*, 256–283. [\[CrossRef\]](#)
25. Oh, S.; Ryu, B.; Shin, Y. EPC signaling load impact over S1 and X2 handover on LTE-Advanced system. In Proceedings of the 2013 Third World Congress on Information and Communication Technologies (WICT 2013), Hanoi, Vietnam, 15–18 December 2013; pp. 183–188. [\[CrossRef\]](#)
26. Rumelhart, D.E.; Smolensky, P.; McClelland, J.L.; Hinton, G.E. Schemata and Sequential Thought Processes in PDP Models. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 2: Psychological and Biological Models*; MIT Press: Cambridge, MA, USA, 1986; pp. 7–57.
27. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Chung, J.; Gülçehre, Ç.; Cho, K.; Bengio, Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv* **2014**, arXiv:1412.3555.
29. Song, C.; Qu, Z.; Blumm, N.; Barabási, A.L. Limits of Predictability in Human Mobility. *Science* **2010**, *327*, 1018–1021. [\[CrossRef\]](#) [\[PubMed\]](#)