Energy-Efficient Neural Network-Based Handover Prediction in 6G Networks

Fatma Laribi, Chaima Ghribi, and Laurent Ciavaglia
Nokia, Massy, France
{*fatma.laribi,chaima.ghribi,laurent.ciavaglia*}@*nokia.com*

***Abstract*—This paper presents a novel approach for balancing energy efficiency and QoS in the context of neural network-based predictive handover management. Our solution encompasses a two-step approach. Initially, we collect data on the energy consumption of the neural network model. Subsequently, a grid search algorithm is deployed to ascertain optimal hyper-parameter thresholds. This process effectively optimizes the trade-off between energy efficiency and QoS. Our experimental results showcase substantial energy savings without significantly compromising QoS. To the best of our knowledge, this paper is the first study exploring balance between energy efficiency and the performance of neural-based predictive handover.**

***Index Terms*—Predictive Handover, Neural Network, Energy Efficiency, Qos, Hyper-parameter Tuning.**

I. Introduction

Mobility management and predictive handover are increasingly challenging in the landscape of 5G and beyond networks, characterized by constantly moving user equipment (UEs) and small cell deployments. Effectively handling these issues is essential to ensure a high quality of service (QoS) and quality of experience (QoE) for UEs.

AI and machine learning techniques have shown potential in enhancing handover decisions in 5G and future networks. While some studies have applied Reinforcement Learning (RL) methods [**?**], [2], [12], [13], [15], it's important to note that RL comes with computational demands and a lack of interpretability. Some other studies have explored the use of LSTM neural networks for predictive handovers, which efficiently process sequential data and offer interpretable predictions [3], [5]–[7], [9], [14]. These LSTM-based approaches have exhibited improved efficiency in optimizing UE mobility by determining the ideal timing and target cells. However, these studies have predominantly focused on performance and QoS aspects, leaving a notable gap in the examination of the energy efficiency of the models themselves. Our work distinguishes itself by proposing a solution to balance QoS and energy efficiency within LSTM models for predictive handover management. This novel approach aims to address the crucial equilibrium between network performance and energy conservation, addressing a significant research gap.

In this paper, we address the critical challenge of striking a balance between QoS and energy efficiency in predictive handover management. Our proposed solution involves a two-step approach. First, we gather data on the energy consumption of the neural network model over time. Subsequently, we employ a grid search algorithm to identify optimal hyper-parameter thresholds, effectively optimizing the trade-off between energy efficiency and QoS. Our experimental results demonstrate significant energy savings without a substantial loss in QoS. To the best of our knowledge, this study is the first in the literature to tackle the issue of finding equilibrium between energy efficiency and the performance of neural-based predictive handover.

The remainder of this paper is structured as follows. Section II presents a brief overview of the related work. Section III describes our proposed solution . Section IV shows details of our experimental tested and shows experimental results followed by conclusions in Section V.

II. Related work

AI/ML techniques can enhance the efficiency and effectiveness of handover decisions in 5G and beyond networks. Several studies have leveraged machine learning (ML) techniques to enhance the efficiency of HO by learning the optimal time and destination for handovers in 5G radio networks. One trend was the application of Reinforcement Learning (RL) methods. Some RL-based approaches have been proposed in [**?**], [2], [12], [13], [15] and were involving training models through online interaction with the network environment, obviating the necessity for pre-existing training datasets. While RL is efficient in terms of adaptability and real-time decision-making, it often presents challenges due to its high computational training demands. Furthermore, RL models tend to lack interpretability, rendering it difficult to grasp the underlying reasoning for their decisions.

On the other hand, some works have been presented in the litterature to solve predictive handover using LSTM neural networks that are efficient for processing sequential data, offering a data-dependent prediction that leverages historical information. Those approaches are more interpretable and adept at capturing long-term dependencies. Authors in [7] proposed a predictive handover (HO) approach using LSTM neural networks, to optimize UE mobility based on real-time measurements and predefined performance objectives. Their study involved developing a predictive classifier that determines the optimal timing and target cell for handovers, aiming to minimize the total network interruption time (MIT). Authors in [6], conducted a comparative study of predictive models for mobility management, using real data collected from a live network, comprising approximately 8 million daily session data records from tracked areas. Their findings indicate that the LSTM model outperforms baseline, linear, and CNN-based models in accurately predicting cellular mobility management traffic volumes. Authors in [9], proposed a "Deep-Mobility" model using deep learning neural networks (DLNN) like LSTM or RNN. This model uses network KPIs to monitor RF signal conditions and system-level inputs, allowing it to make informed handover decisions for network mobility management. Other works like [3], [5] and [14] have proposed also some LSTM based approach. The prior cited studies focus on examining performance and quality of service (QoS) aspects. However, a noticeable gap exists when it comes to evaluating the energy efficiency aspects of the adopted models. While performance and QoS are undoubtedly critical, the importance of addressing energy efficiency in the context of network management cannot be overstated.

While some research studies like [4], [8], [10], [11] have explored the integration of Quality of Service (QoS) and energy efficiency considerations in the context of handover, these efforts have focused on energy efficiency optimization techniques. These optimizations have hinged on leveraging Long Short-Term Memory (LSTM) to capture temporal correlations in channel data, and employing a feedforward network to make informed on/off mode decisions for base stations (BSs). Theses approaches can lead to substantial reductions in cumulative energy consumption through resource optimization and policy-driven control of BS states. However, it's noteworthy that the prior studies have not examined the intrinsic energy efficiency of the LSTM model itself nor addressed the need of striking a balance between QoS and energy consumption in LSTM-based predictive handover.

The distinguishing feature of our work lies in its proposal of a solution to harmonize QoS and energy efficiency considerations within LSTM models, offering a novel approach to address this crucial equilibrium.

III. Energy-Efficient Neural Network-Based Handover Prediction

A. Neural Network-Based Predictive Handover

Our study is built upon the basis of the Neural Network-Based Predictive Handover model presented in [7] and that allows the mobility to be dynamically optimized for each UE and situation while considering the predicted radio conditions. The model is based on predictive classifier that takes UE measurements and determines the optimal point in time and the target cell to HO, in order to minimize the total Mobility Interruption Time (MIT).

As depicted in Figure 1, the model uses UE Reference Signal Received Power (RSRP) data from K designated cells as input and learns to predict the probability that a given cell will have the best RSRP by a future time instance J. Essentially, the model learns a K-class classification using an LSTM neural network.

The aim was to take the initial model as a starting point for the study and enhance it by striking a balance between the model's energy consumption and its accuracy/precision. Details regarding the model's energy monitoring and optimization can be found in the following subsections.

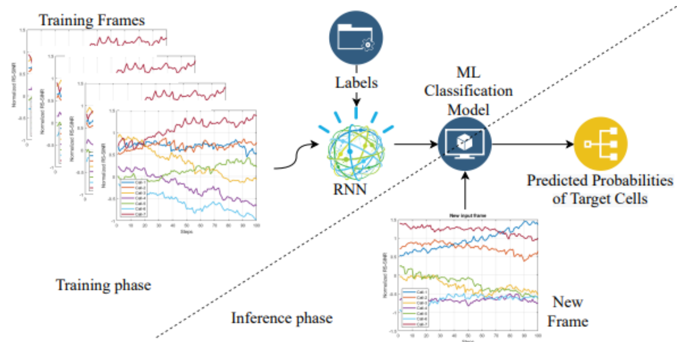

Fig. 1: Predictive Handover [7]

B. Energy Monitoring

In order to monitor and analyze the energy consumption of our predictive handover model, we seamlessly integrated CodeCarbon [1], an open-source Python library. This tool tracks energy impact within Python code and provides detailed measurements of power and energy consumption, as well as valuable insights into the model's environmental footprint. This integration with CodeCarbon plays a pivotal role in our quest to strike the vital balance between energy efficiency and high-quality service delivery in the realm of predictive handover management.

C. Hyper-parameters Tuning Optimization

We used an adapted grid search algorithm for hyperparameter tuning to balance accuracy and energy efficiency.

As depicted in Algorithm 1, the algorithm begins with an initialization step, defining a grid of hyperparameters for exploration, including variables which impact both accuracy and energy efficiency (e.g. learning rates, batch sizes..). Training and evaluation are carried out for each set of hyperparameters, using a subset of training data and considering both accuracy and energy consumption. Energy measurement are provided via CodeCarbon, that monitor the model's energy usage during training. A scoring function f_s was proposed to balance accuracy and energy consumption.

$$f_s(m) = \alpha \cdot \mathrm{A}(m) - \beta \cdot \mathrm{E}(m) \tag{1}$$

Where for a model m, α is a weight parameter for Accuracy, and β is a weight parameter for Energy consumption that allow to adjust the trade-off between accuracy and energy consumption. We consider $\alpha = \beta = 0.5$ to balance accuracy and energy consumption equally. $A(m)$ is the accuracy of the model and $E(m)$ represents the energy consumption reported by Codecarbon tool.
Then, the algorithm iteratively searches for the best hyper-parameter set that achieves the desired trade-off between accuracy and energy efficiency.

Algorithm 1 Hyperparameter Tuning with Energy-Accuracy Balance

Initialization: Start with a set of hyperparameters and define a grid search space.
while not converged **do**
  for each combination of hyperparameters in the grid **do**
    - Train a model with the given hyperparameters.
    - Measure accuracy and energy consumption during training.
    - Calculate a score that balances accuracy and energy.
  end for
  Select the hyperparameters that result in the best score.
end while=0

IV. Experimental Results

A. Experimental Setup

All experiments conducted in this study were carried out on our private Cloud platform to address the predictive handover use case, employing a Long-Short Term Memory model and to have initial insights and evaluations of our novel Hyperparameter Tuning Optimization approach. As previously mentioned, we utilized the CodeCarbon library to monitor the energy efficiency of this model across various parameter configurations.

During the experiments, we have applied our Hyperparameters Tuning Optimization approach while varying the parameters of the predictive handover model. Our objective was not only to fine-tune the model's performance but also to observe the associated energy consumption and CO2 emissions across key tasks, which included data acquisition, model training, and evaluation. Our overarching aim was to balance model performance with sustainability and environmental considerations.

The key hyperparameters that formed the focus of our investigation included:

- **Batch Size:** This parameter governs the number of data points processed in both the forward and backward passes during training.
- **Number of Epochs:** We explored a range of values for this parameter to assess its impact on model convergence and accuracy.
- **Split Ratio:** This pertains to the division of the dataset into training, validation, testing, and the remaining data, each with its unique significance.
- **Hidden Size:** We examined the consequences of altering the size of the hidden layer, a critical intermediary layer positioned between the input layer (receiving raw data) and the output layer (generating final predictions).
- **Learning Rate:** This essential hyperparameter, represented as a positive scalar, was subjected to rigorous scrutiny as it plays a pivotal role in determining the step size during each backward pass of the training process.

In the following section we will delve into an in-depth analysis of the obtained results, emphasizing the significance of balancing model performance with environmental concerns.

B. Results

[To reformulate, I have added the paragraphs as they are for now, and diagrams will also be redone with Python]

In the forthcoming result analysis, our evaluation will be structured in alignment with the hyperparameters previously cited. This approach will enable a focused and comprehensive assessment of how variations in these key parameters influence the performance and energy efficiency of the predictive handover model.

1) Batch Size: Beginning with the exploration of batch size, we consistently observed high accuracy levels within the range of 5 to around 50, while simultaneously noting a decrease in energy consumption. These findings highlight a compelling synergy where a batch size of 50 strikes an attractive balance between energy efficiency and accuracy.

Larger batch sizes can enhance GPU utilization efficiency and potentially reduce energy consumption. However, excessively large batch sizes may necessitate increased memory resources, potentially leading to elevated energy consumption. The correlation between batch size and energy consumption exhibits a descending trend as batch size increases, up to a certain threshold. Beyond this point, further increases in batch size result in an upswing in both energy consumption and CO2 emissions, as exemplified when a batch size of 300 is deemed excessive. This nuanced relationship between batch size and energy efficiency underscores the importance of optimizing this parameter to strike the right balance.

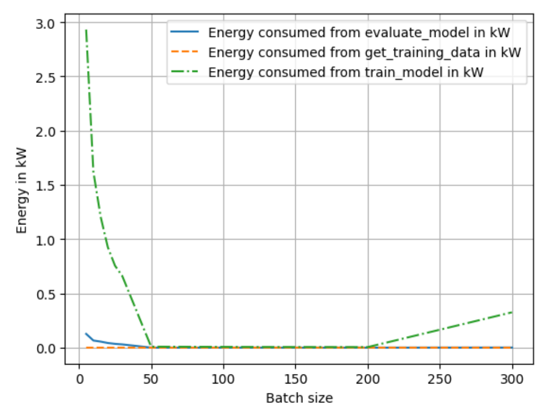

Fig. 2: Energy Consumption Variation with Batch Size

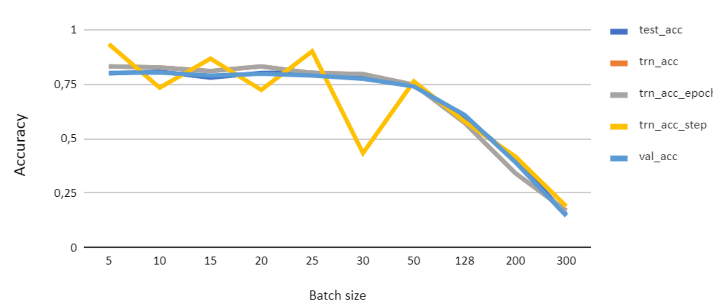

Fig. 3: Accuracy Variation with Batch Size

2) Number of Epochs: While varying the number of epochs, a noticeable trend emerges: as the number of epochs increases, both energy consumption and CO2 emissions also increase. This relationship is driven by longer training times associated with a higher number of epochs, which, in turn, result in heightened energy consumption and emissions. This extension in training time is attributed to the model undergoing a greater number of iterations, necessitating more computations.

In our analysis, we observed distinct phases when examining accuracy and energy/CO2 diagrams. The first phase, spanning from 5 to 50 epochs, demonstrates a significant increase in accuracy, ranging from 60% to 90%, while energy consumption remains minimal. In the second phase, between 50 and 150 epochs, accuracy, though slightly reduced, remains high (exceeding 80%), and energy consumption begins to show a slight increase, although it is still relatively low. The third phase, from 150 to 200 epochs, is characterized by a notable surge in energy consumption, while accuracy experiences a slight decline. Notably, after reaching 50 epochs, accuracy exhibits minimal fluctuations, indicating convergence towards the highest achievable accuracy. In light of these findings, the optimal choice for the number of epochs falls within the range of 50 to 150. Within this range, accuracy remains high, and energy consumption is low.

While there are minor accuracy fluctuations towards the end, these differences can be considered negligible. Overall, our observations lead to the conclusion that a good interval for the number of epochs lies between 50 and 150. Within this range, the model maintains a high level of accuracy while keeping energy consumption at a minimum. This highlights the trade-off between the number of epochs, accuracy, and energy consumption.

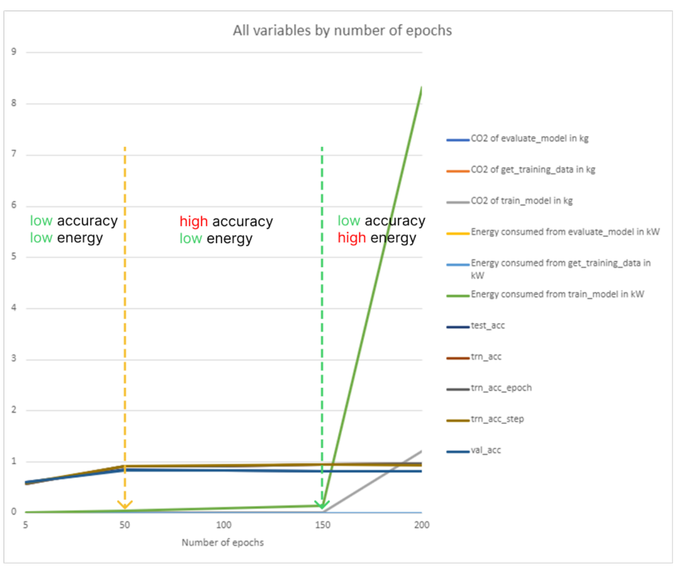

Fig. 4: Accuracy Variation with Number of Epochs

3) Split Ratio: Regarding the split ratio, our observations highlight a significant impact on accuracy, particularly when the training split is substantial. In such cases, where the model has been trained on a larger portion of the data and consequently requires more time, accuracy tends to be notably high. However, it's important to note that this increase in training split does not necessarily guarantee higher accuracy, as it may introduce the risk of over-fitting. Therefore, the selection of the optimal split ratio must consider this balance.

In this context, it appears that a training split of 70% strikes a favorable equilibrium. While it offers high accuracy, it also keeps energy consumption within reasonable limits. This suggests that an excessive training split may lead to unnecessarily high energy consumption without a proportional increase in accuracy, reinforcing the need for a judicious choice of the split ratio.

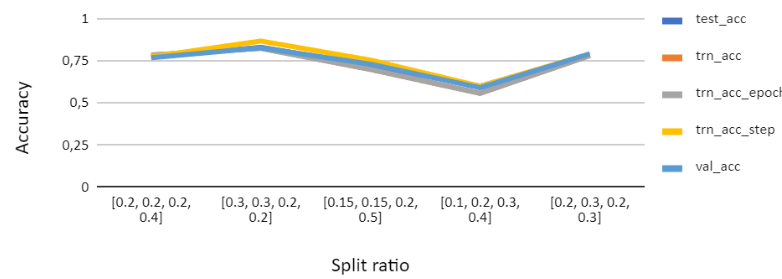

Fig. 5: Accuracy Variation with Split Ratio

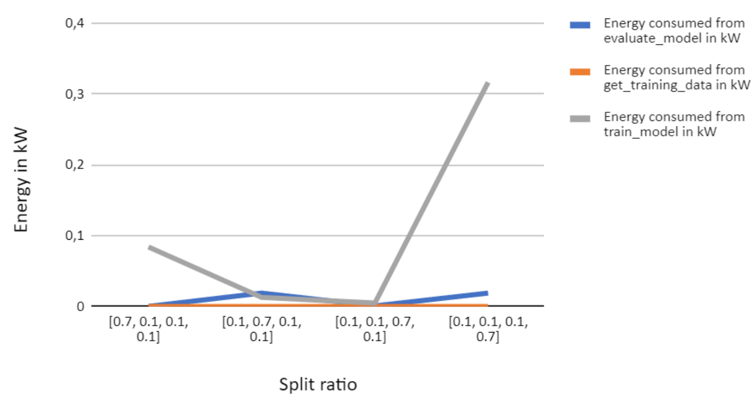

Fig. 6: Energy consumption Variation with Split Ratio

4) Hidden Size: Concerning the hidden size parameter, our analysis reveals a clear relationship: higher hidden sizes correspond to increased accuracy, while energy consumption tends to fluctuate. This observation delineates three distinct phases:

- From a hidden size of 64 to approximately 250: This phase is characterized by low accuracy and low energy consumption.
- Between hidden sizes of 250 and 510: In this phase, accuracy is notably high, and energy consumption remains relatively low.
- From hidden sizes of 520 to 1024: This phase showcases both high accuracy and high energy consumption.

Within this complex interplay, there exists an optimal parameter value that balances high accuracy with reasonable energy consumption. In this particular instance, a hidden size of 512 emerges as a potential ideal parameter value. Here, accuracy exceeds 70%, and energy consumption remains within acceptable limits, making it a favorable choice in achieving the desired equilibrium between accuracy and energy efficiency.

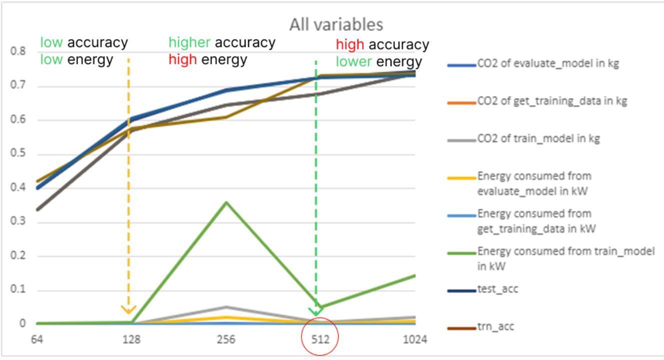

Fig. 7: CO2, Energy and Accuracy Variation with Hidden Size

5) Learning Rate: Concerning the learning rate, higher values tend to accelerate convergence, resulting in reduced training time and lower energy consumption. However, they can also introduce instability, necessitating additional iterations to achieve optimal performance, potentially leading to increased energy consumption. This trade-off is noticeable when examining the case of a learning rate set at 0.001.

In the context of learning rate (lr), it becomes evident that as accuracy improves with higher lr values, a practical approach involves experimenting with elevated values while concurrently evaluating their impact on energy consumption. For instance, with a learning rate of 0.1, the highest value seems to strike an advantageous equilibrium. It delivers both a high level of accuracy and minimizes energy consumption, making it a suitable choice for achieving an optimal balance between accuracy and energy efficiency.

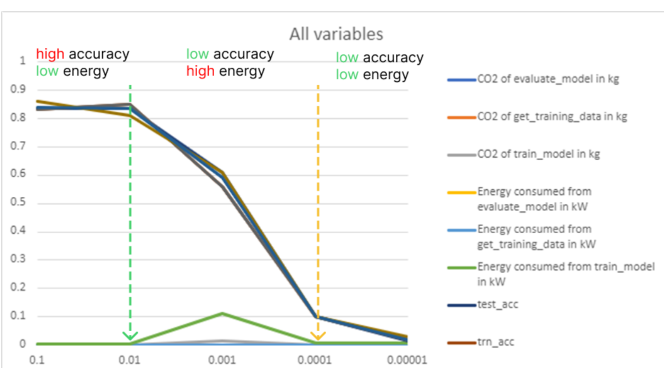

Fig. 8: CO2, Energy and Accuracy Variation with Learning Rate

V. Conclusion and Future work

This paper tackles the problem of balancing Quality of Service (QoS) and energy efficiency in predictive handover management. Our approach involves a two-step process: first, collecting data on the energy consumption of the neural network model over time, and then employing a grid search algorithm to identify optimal hyperparameter settings. This effectively optimizes the trade-off between energy efficiency and QoS. Our experiments demonstrate substantial energy savings without sacrificing QoS, making this the first study in the literature to address the equilibrium between energy efficiency and neural-based predictive handover performance. Our findings emphasize the importance of hyperparameter tuning and its role in achieving a balance between performance and sustainability. Future work will involve a more comprehensive evaluation and further enhancements to our approach.

References

[1] CodeCarbon.io. Codecarbon: Measure and reduce code carbon emissions, Accessed on 2023-10-23.

[2] Younghoon Jang, Syed M. Raza, Hyunseung Choo, and Moonseong Kim. Uavs handover decision using deep reinforcement learning. In *2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, pages 1–4, 2022.

[3] Jaeseong Jeong, Dinand Roeland, Jesper Derehag, Åke Ai Johansson, Venkatesh Umaashankar, Gordon Sun, and Göran Eriksson. Mobility prediction for 5g core networks. *IEEE Communications Standards Magazine*, 5(1):56–61, 2021.

[4] Seungnyun Kim, Junwon Son, and Byonghyo Shim. Energy-efficient ultra-dense network using lstm-based deep neural networks. *IEEE Transactions on Wireless Communications*, 20(7):4702–4715, 2021.

[5] Yuping Lu, Cuntao Zhang, Deen Chen, Wanle Zhang, and Ke Xiong. Handover enhancement in high-speed railway 5g networks: A lstm-based prediction method. In *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–6, 2022.

[6] Lajos Bence Makai and Pal Varga. Predicting mobility management demands of cellular networks based on user behavior. In *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, pages 1–6, 2023.

[7] Ahmed Masri, Teemu Veijalainen, Henrik Martikainen, Stephen Mwanje, Janne Ali-Tolppa, and Márton Kajó. Machine-learning-based predictive handover. In *2021 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, pages 648–652, 2021.

[8] Jihoon Moon, Seungnyun Kim, Hyungyu Ju, and Byonghyo Shim. Energy-efficient user association in mmwave/thz ultra-dense network via multi-agent deep reinforcement learning. *IEEE Transactions on Green Communications and Networking*, 7(2):692–706, 2023.

[9] Rahul Arun Paropkari, Anurag Thantharate, and Cory Beard. Deepmobility: A deep learning approach for an efficient and reliable 5g handover. In *2022 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, pages 244–250, 2022.

[10] DG Rubasinghe, VJ Pussella, MASD Marasinghe, SARKT Silva, and Udaya Dampage. Qos aware energy efficient vertical handover approach in lte/wlan heterogeneous networks. In *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, pages 467–472, 2020.

[11] Chinchaya Sapwatcharasakun, Sunisa Kunarak, and Komkrit Prasertwong. Green energy of handover management in seamless networks. In *2023 International Conference On Cyber Management And Engineering (CyMaEn)*, pages 392–396, 2023.

[12] Yujae Song, Sung Hoon Lim, and Sang-Woon Jeon. Handover decision making for dense hetnets: A reinforcement learning approach. *IEEE Access*, 11:24737–24751, 2023.

[13] Shweta Verma, S Abhirami, Abhinav Kumar, and Sai Dhiraj Amuru. Double deep reinforcement learning assisted handovers in 5g and beyond cellular networks. In *2023 15th International Conference on COMmunication Systems NETworkS (COMSNETS)*, pages 466–470, 2023.

[14] Chujie Wang, Lin Ma, Rongpeng Li, Tariq S. Durrani, and Honggang Zhang. Exploring trajectory prediction through machine learning methods. *IEEE Access*, 7:101441–101452, 2019.

[15] Vijaya Yajnanarayana, Henrik Rydén, and László Hévizi. 5g handover using reinforcement learning. In *2020 IEEE 3rd 5G World Forum (5GWF)*, pages 349–354, 2020.