International Journal of Recent Engineering Science
ISSN: 2349-7157/ https://doi.org/10.14445/23497157/IJRES-V12I6P102

Volume 12 Issue 6, 12-24, Nov-Dec 2025

Original Article

Optimization Model for Load Balancing in 5G Three Tier Heterogenous Networks

Ezeilo Ifeoma Kelechukwu[1], Ifeoma Bernardine Asianuba[2]

[1]*Center for Information and Telecommunication Engineering University of Port Harcourt Rivers State, Nigeria.*
[2]*Department of Electrical/Electronic Engineering, University of Port Harcourt, Rivers State, Nigeria.*

Corresponding Author : windofgreatness@gmail.com

Received: 02 October 2025 Revised: 03 November 2025 Accepted: 19 November 2025 Published: 09 December 2025

Abstract - *Load imbalance is a scenario that occurs in a HetNet where the macrocell is overcongested with loads while leaving the smaller cells underutilized. Over the years, this has remained one of the major issues affecting the quality of service in the 5G network and has continued to attract research interest. In this work, the ability to identify the dynamic characteristics of real-time user behavior that make cell overload a complex problem was established. The methodology required the use of machine learning and simulation to address the challenges of load balancing. The dynamic network data was collected and applied to train a multi-layered neural network algorithm, utilizing dropout regularization and optimization techniques. An overload detection model was formulated and integrated with a vertical handover technique to develop an Intelligent Overload Management (IOM) model in a 5G network. Matlab classification software was used to train the neural network, while the Python programming language was used to integrate the IOM into the 5G HetNet. Simulation results considering varying load factors presented an average latency of 140ms and a throughput of 80%. To validate the model, real-time data from the HetNet was collected during a live walk-test experiment using Open Signal software and then compared with the formulated 5G network with IOM system integration. The results showed an average latency of 139.8ms with IOM as against 148ms in the existing network, presenting a 5.4% improvement in latency reduction.*

Keywords - *Optimization, Heterogeneous network, 5G, Intelligent overload management, Load balancing, Latency, Handover.*

1. Introduction

Load imbalance is a scenario that occurs in a HetNet where the macrocell is overcongested with loads, while the smaller cells remain underutilized.

Massive growth in the use of devices and applications has substantially increased the need for mobile broadband services. Fast and reliable communication in 5G networks has led to challenges with load balancing due to high data usage, varying user demand, resource allocation complexity, impaired Quality of Service, and interference.

Over the years, this has remained one of the major issues affecting the quality of service in the 5G network and has continued to attract research attention.

The invention of the 5G network was tailored towards providing improvement to broadband speed, low latency, massive user connectivity, high wireless transmission and reception connectivity, energy management, and improved efficiency among many other benefits (Ullah et al., 2023). To ensure realization of these benefits, the administrators of the 5G network deployed the use of Radio Network Controllers (RNC) which allows the integration of multiple cells in a heterogeneous format (5GPPP, 2020; Okasaka et al., 2016); however, while this integration of cells has the potential to offer the afore mentioned benefits, Polgar and Varga (2023) revealed that it also has its consequences such as issues of interference, uneven distribution of resources, leading to congestion, call drop, increases energy consumption, handover failure, among others. (Ohaneme et al., 2020; Uguru et al., 2021; Chinedu et al., 2023).

Alireza et al (2019) and Belgium et al (2014) considered other methods to increase system capacity, which are classified into three broad categories: 1. Radio spectrum should be increased, 2. Multiple-Input Multiple-Output antenna transmissions (MIMO) should be used to improve link efficiency, and 3. The radio spectrum should be increased, and Multiple-Input Multiple-Output antenna transmissions (MIMO) should be used to improve link efficiency.

According to Hassan and Fernando (2020), one major reason for these challenges in 5G heterogeneous networks is due to the cells (micro, pico, femto, and macro) having different transmission power, data rates, capabilities, coverage areas, and hence making it a challenging task for the conventional algorithms in the RNC to ensure efficient management of user equipment. Among these resulting problems of 5G HetNet, Shami et al. (2019) revealed that the load balancing problem has continued to gain increased research attention.

Furthermore, new applications for 5G networks and beyond are surely put to work in recent times, thanks to mobile broadband (eMBB), ultra-reliable Low Latency Communication (uRLLC), and massive Machine Type Communications (mMTC). Kalafatidis et al. (2022). Load Balancing (LB) problem according to Suresh et al. (2022) occurs when there is an unfair distribution of resources across cells within the heterogeneous network, thus resulting in the weaker or smaller cells being underutilized, while the bigger cells being overloaded and hence leading to technical, social and economic problems such as poor quality of service, congestion, poor user experience, and overall negative impacts on daily user activities which are dependent on the network (Alzuaidi et al., 2022; Scalise et al., 2024).

To solve this problem, a literature review and a gap were identified, which is the inability to capture the dynamic characteristics of an overloaded 5G network in real time.
Expert consultation, machine learning, and simulation methodologies were applied.

2. Review of Related Works

Hatipoglu et al. (2020) presented the application of a handover-based load balancing algorithm for 5G heterogenous networks. The study presented an algorithm that adapts HO Margin (HOM) and Time to Trigger (TTT) based on user speed and received signal reference power. Additionally, the suggested method distributes the loads evenly among nearby BSs. The suggested method also seeks to lower the ratios of HO Failure (HOF) and HO Ping-Pong (HOPP). The system's implementation results show that the suggested technique reduces the network's BS loads' standard deviation. In HetNets, the suggested approach minimizes overload and accomplishes load balancing. Therefore, there is more than 60% improvement in terms of the HOF, and a 63% more balanced network is achieved with the proposed algorithm in the network.

Basu et al. (2020) presented an adaptive control plane for load balancing in SDN-enabled 5G networks. By appropriately placing the controller and hypervisor entities at H-C planes in accordance with their probable placements, the work offers a thorough adaptive load-balancing technique. The network service latency in the investigation was maintained within an appropriate tolerance limit. By modifying the parametric values in accordance with system requirements, the suggested technique can also be applied to different network topologies, in addition to the case study topology (AT&T North America). It demonstrated how various assessment configurations affect the results of each of the four latency measures. The results of the analysis of the implemented method were not presented in the study. However, future work will target some interesting facts, such as AI (Artificial Intelligence)- driven task offloading between the Hypervisor plane and Control plane for optimal resource utilization.

Hasan et al. (2021) conducted research on Constriction Factor Particle Swarm Optimization (CFPSO) for load balancing and cell association in 5G HetNets. To enhance the throughput of 5G LTE Advanced Heterogeneous Networks (5GLHNs), their study proposes a system that utilizes a load-balancing algorithm and the CFPSO method for cell attachment. The aim of this strategy is to shift the traffic of Macro eNodeBs (MeNB) Users (MUEs) to smaller cells called Home eNodeBs (HeNBs). The proposed technique increases throughput by up to 44.08% compared to the current index-based approach and by 94.20% compared to the current Matching with Minimum Quota (MMQ) strategy, as indicated by the system implementation findings.

Wang (2021) researched the use of a cell clustering optimization algorithm for load balancing in a 5G scenario. In order to verify the optimisation algorithm's ability to balance load, increase transceiver data, optimise resource allocation, and improve performance effects, this work first introduces the existing cell-clustering schemes. It then proposes an optimisation algorithm for cell-clustering based on the 5G cellular network architecture, which combines path loss and load balancing. Finally, the optimisation algorithm is simulated, and the results are compared with the traditional schemes. After adopting 80 micro-base stations, the suggested system achieved a balance degree of 78.48%, according to the system implementation result.

Abbasi et al. (2022) presented an efficient traffic load balancing algorithm for resource optimization in SDN-driven 5G networks. In the context of SDN-driven 5G networks, the study's goal was to achieve efficient network resource allocation while lowering Operating Expenditure (OPEX). The Heuristic Paths Re-computation (HPR) method performs exceptionally well in resource allocation within this framework. It provides two traffic load balancing algorithms, namely Load-Balancing Breadth-First Search (LBB) and Iterative-Deepening Depth-First Search (IDDFS), to improve network scalability and OPEX efficiency in large-scale 5G networks. The Depth-First Search (DFS) method is used by the suggested IDDFS algorithm to prioritize pathways with the maximum bandwidth while limiting the depth of each search. In addition, the IDDFS algorithm guarantees load balancing while drastically lowering space overhead. As a result, in comparison to the LBB algorithm, the suggested IDDFS method improves the efficiency of global optimisation and decreases or even eliminates pointless searches. The results demonstrate that the network throughput for HPR increases from 940 to 2020; for LBB, it increases from 1250 to 2570; and for IDDFS, it increases from 1330 to 3380.

Mahapatra et al. (2022) presented the use of a multi-tier delay-aware load-balancing technique for the Hybrid Cloud Radio Access Network (HC-RAN) 5G architecture. To get beyond the aforementioned restrictions, the multi-tier HC-RAN architecture described in this study takes advantage of the edge computing notion at the Remote Radio Head (RRH). Furthermore, they suggested a Multi-Tier Delay-Aware Load Balancing (MDALB) method to minimise load at the Centralised Baseband Unit (C-BBU0) and front haul. Incoming flow is strategically distributed between RRH and C-BBU using the suggested algorithm, which was executed over the EO based on a load distribution ratio. The system achieved an average packet loss of 1.5% and a packet-to-end latency of 50 packets/milliseconds, according to the system implementation results.

Gures et al. (2023) presented a study on 5G heterogeneous networks load balancing according to the automated weight function. A fifth-generation (5G) coordinated load balancing approach and LTE-A HetNets were presented in this work. The algorithm automatically optimises the Handover Control Parameters (HCP) settings for a given user based on three bounded functions (the UE's speed, the number of Physical Resource Blocks (PRBs) per UE, and the Signal-To-Interference-Plus-Noise Ratio (SINR)) along with their automatic weight levels, Signal Received Power (RSRP) criterion, and the cell load level. It is also suggested to implement a new Handover (HO) protocol that takes into account the pilot signal power, which takes into account the RSRP and the quantity of Physical Resource Blocks (PRBs) per UE. In user association, cells with free PRBs are given preference to offer improved throughput and load balancing. The study's findings show that the suggested load balancing strategy improves network performance measured by throughput, spectral efficiency, load level, and Call-Dropping Ratio (CDR) for a variety of mobile speed circumstances.

The research on load balancing before 5G was simpler, focused on efficiently using available resource blocks, and typically consisted of surveys that reviewed existing methods without introducing advanced algorithms. With the advent of 5G, the need for more dynamic and complex systems pushed researchers towards algorithm-based load balancing methods.

Most reviews and research papers on load balancing in 5G heterogeneous networks (HetNets) have been conducted across various network environments, including Software-Defined Networks (SDNs), Internet of Things (IoT) networks, and others, employing different methodologies to achieve load balancing. As a result, many of these studies did not specifically focus on the unique characteristics of 5G HetNets.

Many papers on load balancing have covered varied network environments and sometimes narrowed their scope to specific components, such as small cells or IoT networks; however, they often do not fully consider the unique challenges and characteristics of 5G HetNets. Consequently, these reviews may not provide comprehensive insights into the 5G-specific load balancing issues faced in a true heterogeneous network setting.

3. Materials and Methods

The open signal analyzer software is an open-source application for network performance monitoring and analysis, deployed on a mobile phone, and used for live data collection of the 5G HetNet network performance. Google Colab was used for the model simulation using the Python programming language, while MATLAB was utilized for the training of machine learning algorithms in the generation of overload detection models. Excel was utilized for data analysis, where a graphical approach was applied. All the software was installed on the laptop, apart from the Open Signal software.

3.1. Research Methodology

The methodology used for this research is a mixed method. This involves expert consultation, machine learning, and a simulation approach. The expert consultation was necessary due to the interdisciplinary nature of this research, which involves consulting domain experts from diverse disciplines, including data science, electronics engineering, and MTN Nigeria, Enugu. Machine learning is the methodology used to develop the proposed solution for overload detection in the 5G network, while simulation is the methodology engaged in the implementation and testing process.

3.2. Research Design

Firstly, a model of the 5G heterogeneous network was developed. The load balancing optimization problem was mathematically formulated to establish the objective function of the research, which is the even distribution of resources across multiple cells and the sustainability of the quality of service.

The Objective Function maximizes the total network capacity while minimizing load imbalance:

$$\max \sum_{\upsilon \in U_{total}} \sum_{\delta \in S} \sum_{r \in R} a_{u,s,r} \;.\; B_{u,s,r} \;.\; \log_2 \left(1 + \frac{P_{u,s,r} \;.h_u}{I_{u,s,r} + N_o} \right) \tag{1}$$

The objective function further minimizes the difference between the load on each tier L_{i} and the average load L. This can be expressed as:

$$\text{Minimize:} \sum_{i=1}^{3} \left| Li - \tilde{L} \right| \tag{2}$$

Where:
L_i is the load on tier I,
$\tilde{L}$is the average load.
$C_{u,s,r}$ is the capacity for user μ on resource block r in subframe ϑ,

- $B_{u,\vartheta,r,}$ is the bandwidth allocated to user μ on subframe ϑ, and resource block *r,*
- $P_{u,\vartheta,r,}$ is the transmit power for user μ on subframe ϑ and resource block *r*,
- H_u is the channel gain for user μ,
  $I_{u,\vartheta,r}$ is the interference experienced by user μon subframe $\vartheta$and resource block *r*,

N_0 is the noise power.

To address this load imbalance when detected, a handover optimization technique is adopted and applied in fusion with the overload detection model. The techniques are integrated on a 5G network gNodeB using MATLAB and Google Colab software, and then rigorously tested under increased user diversity. To validate the model, a real-time experiment was performed on the 5G HetNet, and the live data collected was utilized for model validation after system integration.

In realization of this section, the techniques used are data collection, data processing, machine learning, Artificial Neural Network (ANN), training of ANN, and then model generation, as shown in Figure 1.

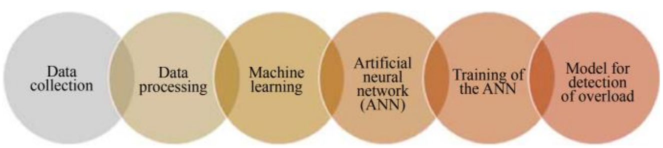


Fig. 1 Sequence diagram for the overload detection model generation

3.3. Data Collection and Preparation

The primary source of data used for this work was collected from MTN Nigeria, Zik Avenue, Enugu. The data modeled the technical specifications of a 5G HetNet, consisting of macro, pico, and micro cells, with an operating frequency of 3500–3600 MHz. The macrocell identification number is T4699, the picocell is T4673, and the microcell is T4607. The micro cell coordinate is at Latitude 6.3740'40" N and Longitude 7.574'90" N, located at Okpara Avenue; the macro cell coordinate is at Latitude 6.43'742" N and Longitude 7.50'775" N and located at Ogui Road; and the pico cell coordinate is at 6.16'27" N and 7.34'14" E and also located at Ogui Road, 500 meters from the macro cell. In addition, live data from this HetNet was collected through real-life experiments and was used for model validation.

The secondary data source was collected from the Kaggle repository, which provided the data for the 5G network overloaded with traffic. The attributes considered for data collection are load factor, throughput, latency, and packet loss. The collected data were prepared using the mean imputation technique (Pandit, 2023). The imputation process was applied to search for missing values and replace them with the predicted mean-related values within the same column.

3.4. Machine Learning Model (Artificial Neural Network)

The machine learning model adopted in this work is the ANN. This ANN is a biologically inspired network of neurons interconnected to form a feed-forward neural network and was utilized in this work. Figure 2 presents the architectural model of a single neuron, which is made of the input Xn, weight w, n bias function, transfer function, activation function, and the output Y. The multi-layered neural network architecture in Figure 2 was trained to generate a model for the detection of overload in a cell.

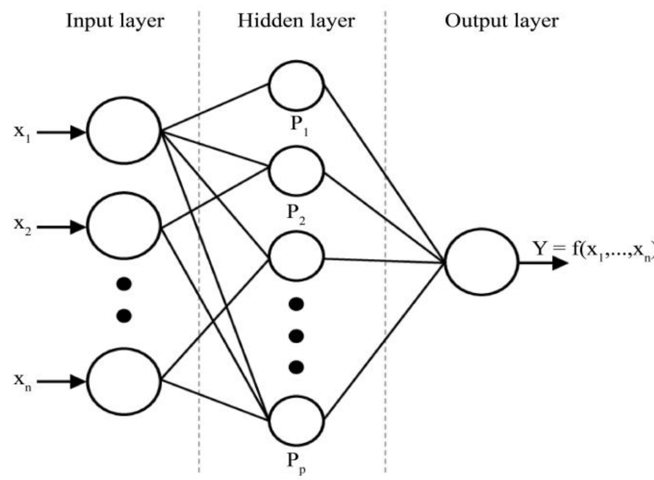


Fig. 2 Architecture of the Multilayered Neural Network (MLN)

Parameters, such as weight, learning rate, and bias, are used to generate the desired model. Prior to the training process, the data was automatically divided into training, testing, and validation by the training application software, which is the neural network toolbox. While the training ensures that the neurons learn the data features, the test and validation sets are applied to ensure the model's performance is okay. To solve the problem of over-fitting, the dropout regularization approach was applied. This dropout ensures that the neurons that have learnt the data are temporarily switched off while allowing other neurons to also learn. The algorithm for overload detection generation is presented as Algorithm 1.

Algorithm 1: Overload model generation with MLN training

1. Start
2. Load dataset
3. Load the neural network model
4. Split data into training, test, and validation sets
5. Load data into the neural network
6. Neural network configuration
7. Set hidden layers, activation function
8. Apply optimization and regularization techniques

9. Train a neural network
10. Evaluate and validate results
11. For best performance
12. Generate a model for overload detection
13. End

Algorithm 2 presents the performance of the overload detection model.

Algorithm 2: Overload detection model
1. Start
2. Initialize network radio controller
3. Allow the incoming packet to the network cell
4. Packet identification
5. Load MLN overload detection model
6. While the packet is fed to a trained MLN
7. Do classification
8. For overload detected
9. Activate the handover model
10. Else
11. Return to step (3)
12. End for
13. End while
14. End

3.5. Handover Optimization Technique

For the load distribution process, the mobile adaptive handover optimization technique in Ogili and Onuigbo (2023) was adapted. The approach is a vertical handover process that first collects the population of available cells within the heterogeneous network and then considers Reference Signal Receive Power (RSRP) and Reference Signal Receive Quality (RSRQ), which are two main parameters considered for the handover process. The flow chart of the process is presented in Figure 3.

Figure 3 presents the flow chart of the handover process when the overload was detected by the intelligent load detection model. To distribute the load, the handover process was initiated by first gathering the population of cells within the Hetnet, then considering the RSRP(-90dbm) and RSRQ(-12db).

The best cell was determined, and the cell channel information was collected and used for the handover request. When the handover request is acknowledged by the radio network controller of the cell, the switching path is configured, and then the user is detached from the main cell and transferred to the selected cell for the handover process.

3.6. System Integration of the Intelligent Overload Management (IOM) Approach

In the overload management approach, the neural network-generated overload detection model and the handover model were integrated as the intelligent optimization approach for overload management, as shown in Figure 4.

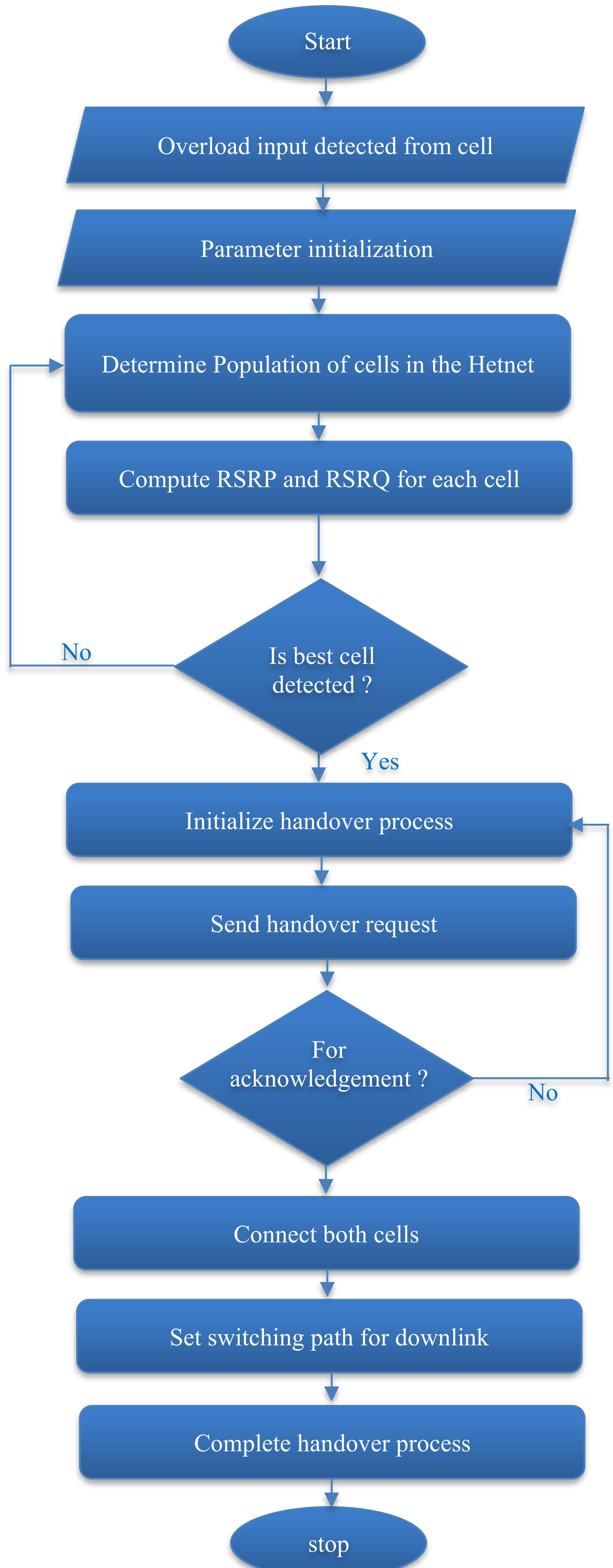


Fig. 3 Flow chart of the handover process

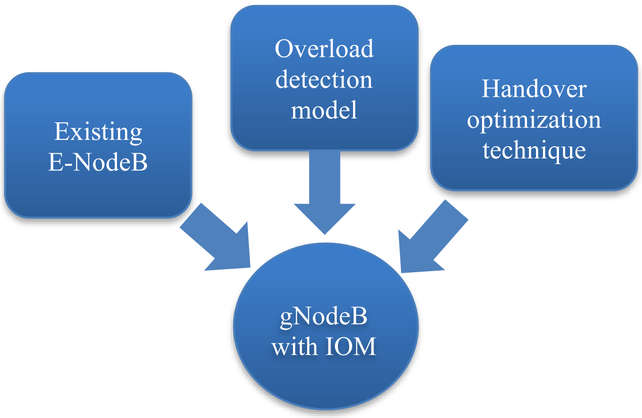


Fig. 4 Integrated gNodeB with IOM

Figure 4 describes a system integration that showcases the two models integrated with the existing gNodeB as the new system for load balancing in a 5G network. The architectural model of the integrated system in a 5G HetNet is presented in Figure 5.

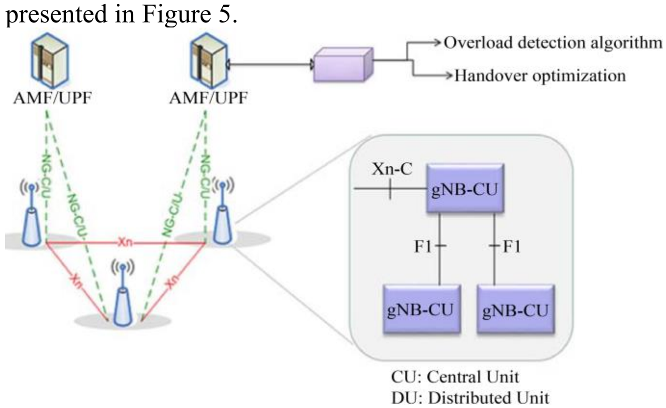


Fig. 5 Architectural model of the integrated IOM on 5G HetNet

Figure 5 shows the case of the integrated IOM on the 5G HetNet. In the diagram, the Access and Mobility Management Function (AMF) manages the initial connection setup, authentication, and mobility management between cells. Inside this AMF is the IOM model developed with the overload detection model and handover techniques, which allows for continued mobility management and load balancing through real-time monitoring of the network performance for overload and then applying the handover techniques for load balancing. The (UPF), on the other hand, is used for traffic handling between user equipment and the Packet Data Network (PDN). The Next Generation Core/User Plan (NG-C/U) provides the core functionalities required, including mobility management and session management. The interface between the gNodeB is presented as Xn, and it facilitates coordination of gNodeB-like handover, radio resource management, and mobility management. The Distribution Unit (DU) provides support for lower-level protocols, such as the MAC and physical layers, while the Control Unit (CU) supports higher-layer protocols.

3.7. Implementation of the New 5G Network with an Intelligent Load Balancing Model

Implementing the new 5G network, which incorporates an overload detection algorithm and an adapted handover technique, was performed using the MATLAB programming language. First, the intelligent load balancing model was conceptualized for real-time monitoring and detection of overload on the HetNet, and it was implemented using statistics, data processing, and a machine learning toolbox in MATLAB. The processed data was imported into the MATLAB classification application software, and then the neural network algorithm was selected and trained using the optimization toolbox, which was configured with the optimization algorithm for neural network training. During the training, the model was evaluated and deployed as software, which was integrated into a 5G HetNet with handover optimization techniques using the Python programming language. To simulate the network, data collected to define the network specifications, as reported, was used to configure it and then tested through simulation at varying load factors. The simulation parameters are reported in Table 1.

Table 1. Simulation Parameters

Cell layout	**Hexagonal grid, 3 sector per site**	**1 Pico per site**	**1 Micro per site**
Carrier frequency	3.5GHz	3.6GHz	3.6GHz
Cell radius	289m (1000m)	-	-
Transmitting Antenna height	32m	10m	15m
Transmitting antenna power	14dBi	4dBi	7dBi
User equipment distribution	15deg	10deg	10deg
Scheduling algorithm	30UE per sector, 2/3 clustered distribution		
Link adaptation	QPSK TO 1024-QAM (25 MCS indexes)		
Traffic model	Full buffer	Number of users	300
MIMO	3*2SU-MIMO	Data type	Uplinks

4. Results and Discussions

The MLN training process during the generation of the overload detection model presented very high efficiency. It also reports the simulation of the IOM on the 5G HetNet. In addition, a real-life experiment was conducted on the existing

5G HetNet, and the data collected was used for the system integration, and all results were reported and analyzed. Finally, the validation results through comparative analysis were also reported and discussed.

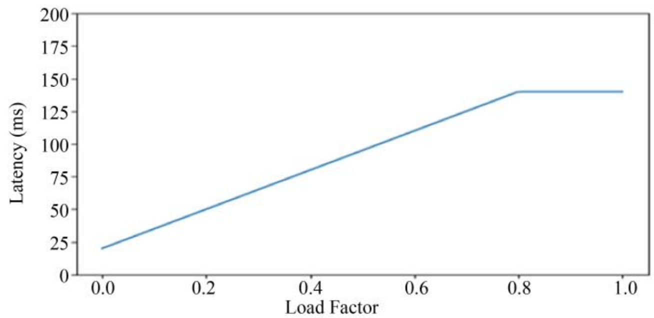


Fig. 6 Result of latency with IOM

Figure 6 displayed the IOM performance on the HetNet during varying load factors. Initially, the latency of the network started to increase as the number of users increased, which is normal. However, as the load factor reaches 0.75, which is the set threshold for overload detection on the network, the intelligent model developed with MLN, as showcased in the algorithm, was triggered, the overload problem was classified, and the handover model activated, which considers the RSRP and RSRQ of neighboring cells to distribute the load and ensure the stability of the network. This was the reason for the steady state experienced by the latency after the 0.8 load factor, thus showing a redistribution of incoming load to control the latency on the network. Overall, it was deduced that the latency never exceeds 150 ms, despite the overload on the network. This value is very good according to the Nigerian Communication Commission's (NCC) standard for quality of service in the 5G network. In another test result, the network's throughput performance was evaluated, considering diverse load utilization factors that model the dynamic characteristics of users' data upload sizes against the overall amount delivered, as reported in Figure 7.

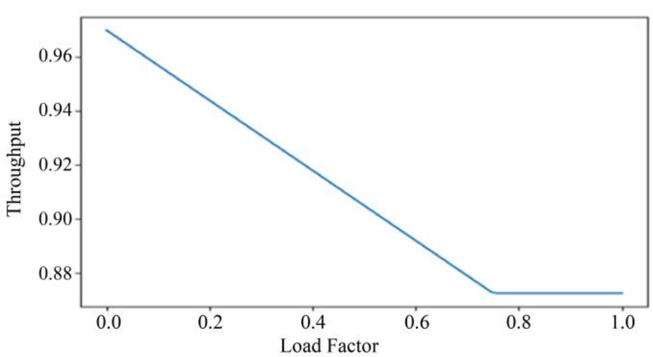


Fig. 7 Throughput result with IOM

Figure 7 reports the throughput performance of the network considering varying load utilization, which is a parameter that models the dynamic characteristics of user and cell behavior during operation. From the result, it was observed that when the load factor kept increasing due to user clustering on the macro cell, the throughput kept increasing too; however, at a 0.75 load factor, the throughput started to remain constant even as the load factor increased. What happens here is that when the load factor reaches 0.75, the intelligent overload detection model is able to classify overload signs on the network and then trigger the handover model, which distributes the load to other smaller cells. So while the load keeps increasing, it is channeled directly to other cells for management while maintaining the quality of service in the main cell. Overall, it was deduced from the graphs that the throughput never exceeds a factor of 0.85, which is equal to 85% throughput on the network, which is in accordance with the NCC standard for 5G as very good.

4.1. System Integration of the IOM

The IOM was integrated into the 5G network, utilizing live data collected from the network through open signal analyzer software. The software was used to collect live network information on the case study area, considering data upload, time, and latency. The data was collected for 30 minutes during a walk test and spanned different times of the day (11:27 a.m. to 4:27 p.m.). Figure 8 shows the software and location of the data collection.

Figure 8 shows the software and location of the data collection.

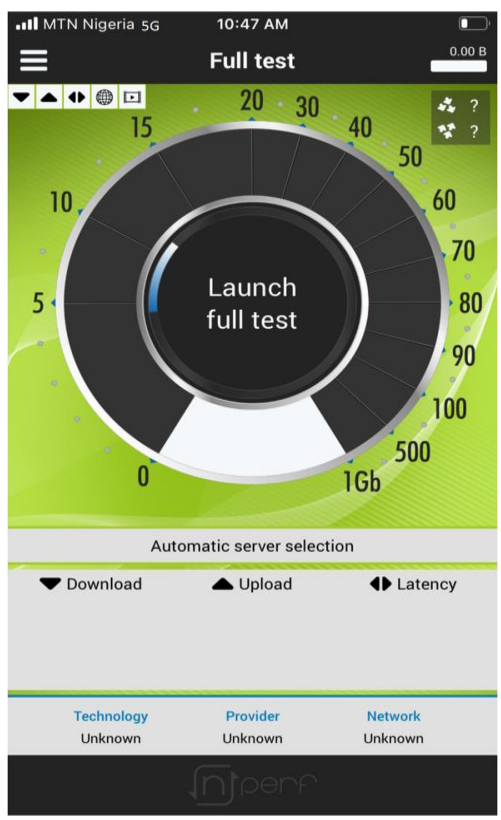


Fig. 8 Network testing

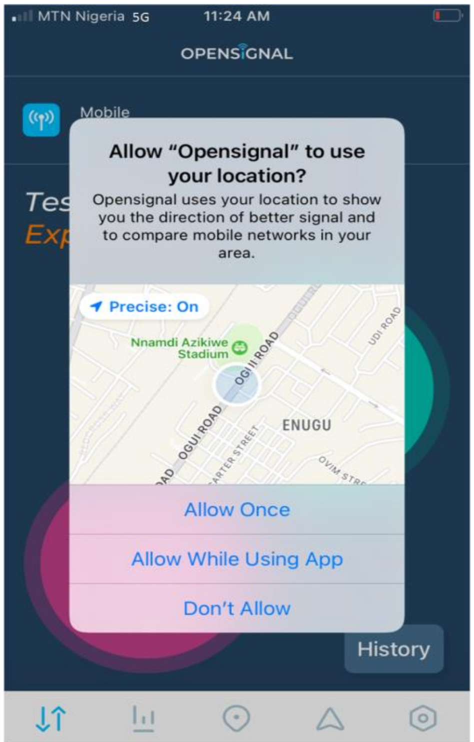
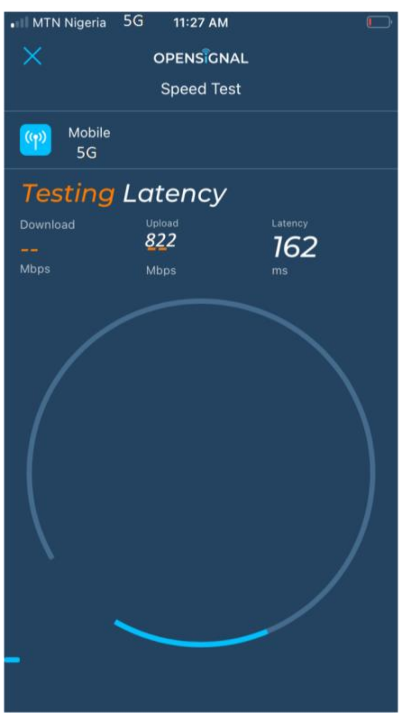
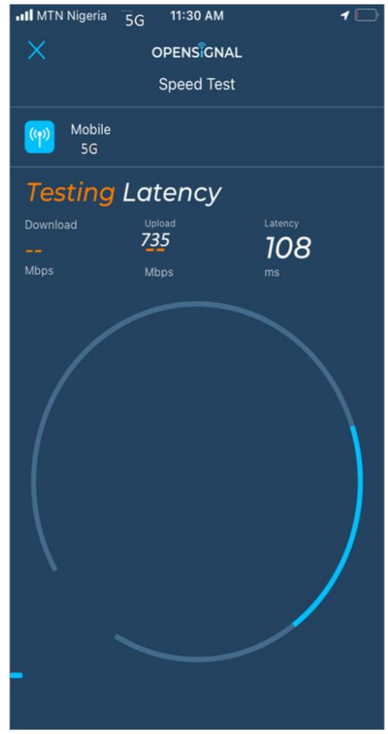


Fig. 9 Location for test

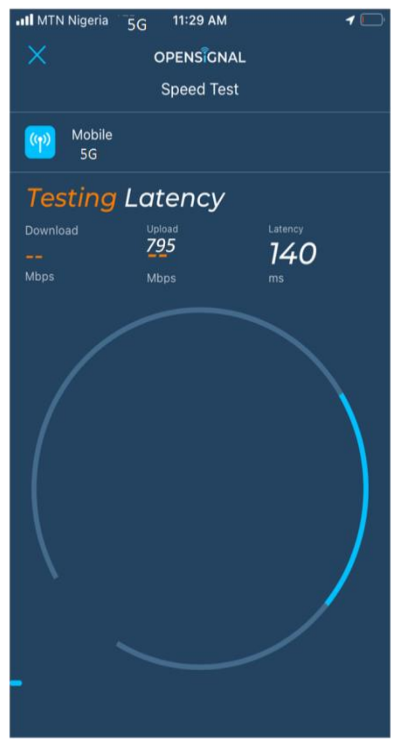
Figure 8 presents the software deployment, while Figure 9 shows the user location during the test process. In running the test, the HetNet performance during uplink was considered at varying times, and the latency was measured as evident in Figure 10 and recorded.



(a)



(b)



(c)

Fig. 10 Walk test latency results of the 5G HetNet

Figure 10 (a, b, and c) respectively showcase the walk test results collected from the 5G HetNet at different locations and varying times of the day, spanning from 11:27 a.m. to 4:27 p.m. The result of the data collection is reported in Table 2.

Table 2. Result of the walk test (Field work) on the 5G HetNet

Time (mm: ss)	Data upload (mb/s)	Load factor	Latency (ms) without IOM
11:27	822	0.802734	162
11:28	801	0.782227	163
11:29	795	0.776367	140
11:30	735	0.717773	108
11:31	603	0.588867	118
11:32	611	0.59668	123
11:33	620	0.605469	126
11:34	621	0.606445	127
11:35	584	0.570313	129
11:36	532	0.519531	105
14:27	930	0.908203	196
14:28	928	0.90625	194
14:29	919	0.897461	190
14:30	908	0.886719	187
14:31	802	0.783203	164
14:32	902	0.880859	180
14:33	908	0.886719	187
14:34	892	0.871094	191
14:35	902	0.880859	196
14:36	970	0.947266	234
16:27	675	0.65918	148
16:28	745	0.727539	157
16:29	745	0.727539	157
16:30	404	0.394531	79
16:31	435	0.424805	85
16:32	643	0.62793	129
16:33	575	0.561523	107
16:34	433	0.422852	89
16:35	746	0.728516	149
16:36	625	0.610352	124
Average	727.0333	0.709994	148.1333

Table 2 presents the results of the data collected from the walk test. A notable observation from the data is that, on average, when 727.0333 MB of data is transmitted per second, the load factor reported is 0.71, and the average latency is 1.45 ms. This result implied that overall the network performance is good; however, it was observed that at certain times of the day, especially from 14:24 to 14:36, the network recorded a very high load factor, which indicated an increase user activities from the cell, and the latency also increased, with the highest latency reporting 234 ms at the peak congestion period of 14:36 p.m. Due to the reactive load-balancing mechanism on the network, it was unable to respond to the overload and manage the poor latency results. The implication is that at this latency, quality of service or delay-sensitive packets, such as voice and HTTP, will be affected, thereby impacting the user experience and prompting the need for optimization and load balancing across the other cells of the network. The next result presents the system integration of the IOM on the 5G HetNet. The integrated system performance was determined for each load factor using real data collected from the 5G network walk test, as well as the reference simulation results.

Table 3 showcases the result of the integrated IOM on the 5G HetNet. From the result, it was observed that when an average packet of 727 was transmitted on the network, the latency was reported at 140 ms. These results are very good, as they imply a good quality of service according to the NCC standard for 5G network analysis. More so, it was observed that from 14:27 to 14:36, the latency appears fairly constant, despite the increased load factor. The reason was that during the network operation, the intelligent overload detection model continuously monitored for overload, and when it was detected, the traffic was redistributed to other cells using handover techniques adapted.

Table 3. Result of system integration

Time (mm: ss)	Data upload (mb/s)	Load factor	Latency with integrated IOM
11:27	822	0.802734	160.5468
11:28	801	0.782227	156.4454
11:29	795	0.776367	135.2734
11:30	735	0.717773	103.5546
11:31	603	0.588867	117.7734
11:32	611	0.59668	119.336
11:33	620	0.605469	121.0938
11:34	621	0.606445	121.289
11:35	584	0.570313	114.0626
11:36	532	0.519531	103.9062
14:27	930	0.908203	181.6406
14:28	928	0.90625	181.25
14:29	919	0.897461	179.4922
14:30	908	0.886719	177.3438
14:31	802	0.783203	156.6406
14:32	902	0.880859	176.1718
14:33	908	0.886719	177.3438
14:34	892	0.871094	174.2188
14:35	902	0.880859	176.1718
14:36	970	0.947266	189.4532
16:27	675	0.65918	131.836
16:28	745	0.727539	145.5078
16:29	745	0.727539	145.5078
16:30	404	0.394531	78.9062
16:31	435	0.424805	84.961
16:32	643	0.62793	125.586
16:33	575	0.561523	112.3046
16:34	433	0.422852	84.5704
16:35	746	0.728516	145.7032
16:36	625	0.610352	122.0704
Average	727.0333	0.709994	139.9987

4.2. Validation of the IOM-Based Load Balancing Model

To validate the new IOM model, comparative analysis was performed, considering the performance of the 5G HetNet with IOM and the network performance without IOM. Table 4 presents the comparative results.

Table 4. Comparative analysis of the 5G HetNet

Time (mm: ss)	Data upload (mb/s)	Load factor	Latency (ms) without IOM	Latency with integrated IOM
11:27	822	0.802734	162	160.5468
11:28	801	0.782227	163	156.4454
11:29	795	0.776367	140	135.2734
11:30	735	0.717773	108	103.5546
11:31	603	0.588867	118	117.7734
11:32	611	0.59668	123	119.336
11:33	620	0.605469	126	121.0938
11:34	621	0.606445	127	121.289
11:35	584	0.570313	129	114.0626
11:36	532	0.519531	105	103.9062
14:27	930	0.908203	196	181.6406
14:28	928	0.90625	194	181.25

14:29	919	0.897461	190	179.4922
14:30	908	0.886719	187	177.3438
14:31	802	0.783203	164	156.6406
14:32	902	0.880859	180	176.1718
14:33	908	0.886719	187	177.3438
14:34	892	0.871094	191	174.2188
14:35	902	0.880859	196	176.1718
14:36	970	0.947266	234	189.4532
16:27	675	0.65918	148	131.836
16:28	745	0.727539	157	145.5078
16:29	745	0.727539	157	145.5078
16:30	404	0.394531	79	78.9062
16:31	435	0.424805	85	84.961
16:32	643	0.62793	129	125.586
16:33	575	0.561523	107	112.3046
16:34	433	0.422852	89	84.5704
16:35	746	0.728516	149	145.7032
16:36	625	0.610352	124	122.0704
Average	727.0333	0.709994	148.1333	139.9987

Table 4 presents the comparative analysis of the 5G HetNet with and without IOM. From the result, it was observed that on average, when 727MB of data was uploaded per second, the average latency with IOM was 140 ms, while the average latency without IOM was 148 ms, thus giving a percentage reduction of 5.41% with IOM. Furthermore, a graphical analysis was performed in Figure 11 to visually show the latency of the two networks at different times of the day.

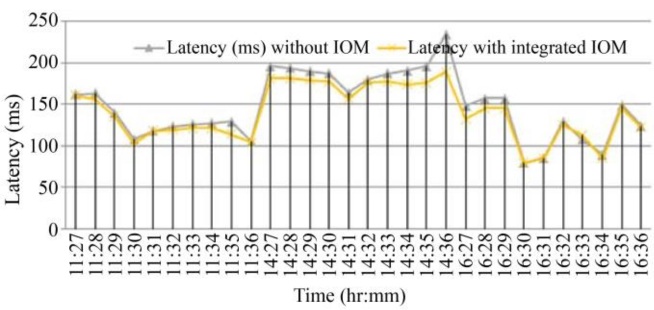


Fig. 11 Comparative latency analysis

Figure 11 showcases the comparative latency of the results of the two networks, one with IOM and the other without IOM. The results revealed that initially, both network latency was very good due to the low utilization factor; however, during the afternoon, precisely around 14:00–14:30, it was observed that the latency increased without IOM, with a peak value of 234 ms occurring at exactly 14:36, while at the same time the latency of the network with IOM was reported at 189 ms. The reasons for the reduced latency on the network with IOM were that the overload detection model was able to detect the pattern leading to the increase in latency and immediately activate the handover mechanism, which redistributes the load to other cells around, considering the cell with the best RSRP and RSRQ.

5. Conclusion

The research highlights the need for real-time load balancing in 5G HetNets, addressing gaps in existing studies that overlook the network's dynamic characteristics during load management.

An intelligent optimization approach was presented for overload management in 5G HetNet using a machine learning technique. A model for overload detection in 5G networks was generated and fused with an adapted vertical handover technique to develop an Intelligent Overload Management (IOM) model for overload management in a 5G network, achieving a percentage improvement of 5.4% for latency reduction.

Over the years, the innovation of 5G HetNet has been introduced to tackle communication challenges such as cell congestion, overload, and interference, among others. While these issues have received considerable research attention aimed at resolution, overload remains a persistent challenge due to the unpredictability of user behavior. To confront this challenge, our research harnesses the potency of machine learning techniques and optimization approaches to craft a model capable of real-time monitoring of overload in 5G networks and subsequently addressing its impact on the user experience through a handover process. This highlights the crucial role of leveraging machine learning techniques and optimization approaches in addressing the challenges posed by overload in 5G networks.

Recommendation

I recommend that the security of data packets during the load distribution process, as well as resource allocation, should be considered. More so, deep learning algorithms can also be introduced while developing the model. Lastly, a three-tier HetNet was used in this study; the application of this model in a higher-tier HetNet can be considered.

Funding Statement

The research was funded by the authors, though we got data and support from MTN Nigeria professionals.

References

[1] Kamran Mustafa Abbasi, "Efficient Traffic Load Balancing Algorithms for Resource Optimization in SDN-Driven 5G Networks," pp. 1-20, 2022. [Google Scholar]

[2] Alireza Alizadeh, and Mai Vu, "Multi-Armed Bandit Load Balancing User Association in 5G Cellular HetNets," *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, Taipei, Taiwan, pp. 1-6, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[3] Osamah Thamer Hassan Alzubaidi et al., "Interference Challenges and Management in B5G Network Design: A Comprehensive Review," *Electronics*, vol. 11, no. 18, pp. 1-49, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] Deborsi Basu, Uttam Ghosh, and Raja Datta, "Adaptive Control Plane Load Balancing in vSDN Enabled 5G Network," *arXiv:2007.09789*, pp. 1-4, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[5] Mohammad Riyaz Belgaum et al., "A Systematic Review of Load Balancing Techniques in Software-Defined Networking," *IEEE Access*, vol. 8, pp. 98612-98636, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] Chinedu R. Okpara, Victor E. Idigo, and Chukwunenye S. Okafor, "Comparative Analysis of the Features of a 5G Network Production Dataset: The Machine Learning Approach," *European Journal of Engineering and Technology Research*, vol. 8, no. 2, pp. 52-57, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Emre Gures et al., "Load Balancing in 5G Heterogeneous Networks Based on Automatic Weight Function," *ICT Express*, vol. 9, no. 6, pp. 1019-1025, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Emre Gures et al., "Adaptive Cell Selection Algorithm for Balancing Cell Loads in 5G Heterogeneous Networks," *Alexandria Engineering Journal*, vol. 72, pp. 621-634, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Mohammad Kamrul Hasan et al., "Constriction Factor Particle Swarm Optimization based Load Balancing and Cell Association for 5G Heterogeneous Networks," *Computer Communications*, vol. 180, pp. 328-337, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Noha Hassan, and Xavier Fernando, "An Optimum User Association Algorithm in Heterogeneous 5G Networks Using Standard Deviation of the Load," *Electronics*, vol. 9, no. 9, pp. 1-20, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[11] Abdussamet Hatipoğlu et al., "Handover-based Load Balancing Algorithm for 5G and Beyond Heterogeneous Networks," *12th International Congress on Ultra-Modern Telecommunications and Control Systems and Workshops*, Brno, Czech Republic, pp. 7-12, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] Sarantis Kalafatidis, and Lefteris Mamatas, "Microservices-Adaptive Software-Defined Load Balancing for 5G and Beyond Ecosystems," *IEEE Network*, vol. 36, no. 6, pp. 46-53, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Byomakesh Mahapatra, Ashok Kumar Turuk, and Sarat Kumar Patra, "Multi-Tier Delay-Aware Load Balancing Strategy for 5G HC-RAN Architecture," *Computer Communications*, vol. 187, pp. 144-154, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[14] Byomakesh Mahapatra et al., "Utilization-Aware VB Migration Strategy for Inter-BBU Load Balancing in 5G Cloud Radio Access Networks," *Computer Networks*, vol. 181, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] C.O. Ohaneme, O.S. Edward, and A.N. Isizoh, "Development of Enhanced Dynamic Cell Sectorization Scheme for Improved WCDMA Network Capacity," *Journal of Engineering and Applied Sciences*, vol. 17, no. 1, pp. 263-274, 2020. [Google Scholar] [Publisher Link]

[16] Shozo Okasaka et al., "Proof-of-Concept of a Millimeter-Wave Integrated Heterogeneous Network for 5G Cellular," *Sensors*, vol. 16, no. 9, pp. 1-21, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[17] Zsolt Alfred Polgar, and Mihaly Varga, "Game Theory-Based Load-Balancing Algorithms for Small Cells Wireless Backhaul Connections," *Applied Science*, vol. 13, no. 3, pp. 1-36, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Paul Scalise et al., "A Systematic Survey on 5G and 6G Security Considerations, Challenges, Trends, and Research Areas," *Future Internet*, vol. 16, no. 3, pp. 1-38, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Tareq M. Shami et al., "Load Balancing and Control with Interference Mitigation in 5G Heterogeneous Networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 177, pp. 1-12, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[20] Krishnamoorthy Suresh et al., "Enhanced Metaheuristic Algorithm-Based Load Balancing in a 5G Cloud Radio Access Network," *Electronics*, vol. 11, no. 21, pp. 1-18, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[21] Simon E. Uguru et al., "Enhanced Interference Management Technique for Multi-Cell Multi-Antenna System," *World Academy of Science, Engineering and Technology International Journal of Electronics and Communication Engineering*, vol. 15, no. 11, pp. 376-384, 2021. [Google Scholar]

[22] Yasir Ullah et al., "A Survey on Handover and Mobility Management in 5G HetNets: Current State, Challenges, and Future Directions," *Sensors*, vol. 23, no. 11, pp. 1-36, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[23] Yuhan Wang, "Study of Cell Clustering Optimization Algorithm for Load Balancing in 5G Scenario," *Journal of Physics: Conference Series*, vol. 1955, pp. 1-8, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[24] Yan Wang et al., "Propagation Dynamics of an Epidemic Model with Heterogeneous Control Strategies on Complex Networks," *Symmetry*, vol. 16, no. 2, pp. 1-22, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[25] Ying-Hong Wang, Guo-Rui Huang, and Yi-Chia Tung, "A Handover Prediction Mechanism Based On LTE-A UE History Information," *International Conference on Computer, Information and Telecommunication Systems*, Jeju, Korea, pp. 1-5, 2014. [CrossRef] [Google Scholar] [Publisher Link]