Explorematics
Journal of Innovative
Engineering and Technology

Volume: 06 No: 01 | April -2025
ISSN (Online) 2636 – 590
ISSN (Print) 2636 - 591X

AN INTELLIGENT TRAFFIC DISTRIBUTION MODEL FOR LOAD BALANCING USING ARTIFICIAL NEURAL NETWORK

Nwadike, Ijeoma M.*[1], Abonyi Dorathy O. [1], Nwadike Kenneth I[2]
1 Department of Electrical and Electronic Engineering, Enugu State University of Science and Technology
2 University of Nigeria, Enugu Campus
Author for correspondence: Nwadike I.M; **E-mail**: aijayij@gmail.com

Abstract - This research aims to use artificial neural networks to create an intelligent traffic distribution model for load balancing. The mathematical model of the 5G heterogeneous network was developed to accomplish this goal. To train artificial neural network (ANN) algorithms and create a model for overload detection, data from the network experiencing overload was gathered. The vertical handover procedure was started using the model as a load distribution control mechanism. The model was evaluated using simulation methodology. Accuracy and loss were considered when evaluating the ANN's training results. The accuracy was 0.89 and the loss function was 0.20; following cross-validation, the accuracy increased to 0.995. The outcome of the system integration showed that handover was used to control the load once overload was detected. With average throughput of 80% for macro cells, 85% for microcells, and 86% for pico cells, key Quality of Service (QoS) metrics show notable gains. For macro cells, microcells, and pico cells, the load factor was lowered to 0.44, 0.28, and 0.23, respectively. Received Signal Received Power (RSRP) values recorded were -80 dBm (macro), -85 dBm (micro), and -86 dBm (pico). In conclusion, the proposed Intelligent Traffic Distribution (ITD) model and Sector-based Traffic Control algorithm effectively mitigated network overload, ensuring better QoS and an enhanced user experience in 5G HetNet environments.
Keywords: Artificial neural network, traffic distribution, overload, 5G, quality of service

1. Introduction

According to Chai et al. (2011), load balancing (LB) is a process that ensures an even distribution of carrier resources from user equipment within a HetNet for optimal quality of service. In addition, Ravyaar et al. (2019) define LB as an optimization solution to the problem of load mismatch between users and cell stations within a HetNet, while Raid (2019) posited that LB ensures the supervision of data flow to avoid overload on one cell via the adjustment of the network parameters using the handover process. In Elmosilhy et al. (2019), the methods to solve the problem of cell overload are the active-active and active-passive load balancing methods. The former has better advantages over the latter as it can understand user traffic information and radio conditions before deciding on load balancing, while the latter has more challenges in achieving load balancing as it lacks the intelligence of user awareness and also lacks other intelligence features to solve the problem of load balancing. Due to this reason, many techniques have been developed under the active mode for load balancing, ranging from machine learning (ML) and genetic algorithms to other standard load-balancing algorithms. The use of ML has been employed to achieve load balancing using reinforcement learning, neural networks, support vector machines, and deep learning, among others, and have been able to accurately detect overloading on cells; however, despite their success, the impact of adjacent interference from other cells and congestion in the channel for the load distribution process was not considered (Raid, 2019; Jordi, 2016; Mohammed et al., 2020; Ye et al., 2019). Standard load balancing techniques employ dynamic algorithms like the call admission control algorithm, smart-base station antenna approach, and dynamic channel assignment, among others (Raid, 2019); however, during this process, interference from other cells and also busy channels when utilized for load balancing through the handover process often results in dropped calls, handover failures, and other poor quality of service factors. In other studies (Eliosilhy et al., 2019; He et al., 2016; Leela et al., 2020; Min et al., 2014; Sepehr and Mohammed, 2019; Kavitha et al., 2020), the load distribution problem was identified as an optimization problem and then applied genetic algorithms such as particle swarm optimization, ant colony optimization, and gravitational search algorithm; however, the impact of interference and congestion during load balancing was not considered. To solve this problem, the researcher proposed an intelligent traffic distribution technique that applied a multi-faceted approach to the load balancing process in the 5G HetNet.

2. Methodology

The methodology of the work developed the heterogeneous network model and then collected data from Mobile Telephone Network (MTN) which modeled the historical congestion patterns on the network from 2022 to 2023 and then applied to train neural network algorithm to generate a data model for the detection of congestion. Vertical handover was applied to control the congestion, and then programming in Matlab was applied to implement the work and then the results were validated through comparative analysis.

2.1 Modeling of the Heterogeneous Network

HetNets are used to improve the quality of service in wireless communication networks, especially in congested environments. The idea is to share load from users among the cells and ease traffic on the main cell which is usually the macro cell. In this case, the HetNet considered for the study consists of three cells which are the macro, micro, and pico cells. The macro being the main cell is supported by the two smaller cells to help manage traffic congestion; however, the dominant properties of the macro cell resulted in issues of load balancing which have to be addressed in this research.

The mathematical definition of the HetNet was presented as;

Macro cell: $M = \{M_1, \dots, M_m,\ M_n \dots., M_M)$ (1)

Where $N_m = \{N_1^{ma}, \dots M_b^{ma}, \dots N_{N_{ma}}^{ma}\}$ is the set of the smaller cell supporting the macro cell $\boldsymbol{M_{ma}}$. The equation 2 was used to present the smaller cells within the HetNet as;

Pico cells:

$P = \{P_1, \dots P_f,\ \ P_g \dots P_p\}$

Microcells:

Mx= $\{Mx_1 \dots Mx_f,\ \ Mx_g \dots Mx_p\}$ (2)

Where N_{mx}= $\{N_1^{mx}, \dots N_b^{mx}, \dots N_{N_{mx}}^{mx}\}$ is the neighboring cell of the micro Mx_{mx}; $N_p = \{N_1^p, \dots N_b^p, \dots N_{N_p}^p\}$ is the neighboring cell of the pico P_p. The users operating within the cell coverage of the HetNet for each individual cell are presented in Equation 3;

Macro Users $M_{ma} = \{U^{ma}, \dots U_1^{ma}, \dots U_u^{ma}, \dots U_{U_{ma}}^{ma}\}$

Pico cell users $P_p :\ U^p = \{U_1^p, \dots U_U^p, \dots U_{U_p}^p\}$(3)

Microcell users $Mx_{pi} :\ U^{pi} = \{U_1^{pi}, \dots U_U^{pi}, \dots U_{U_{pi}}^{pi}\}$

$Subframe: S = \{1, \dots s, \dots S\};\ Resourceblocks\ (RB) = R = \{1, \dots r, \dots, R\}$

2.2 The Intelligent Traffic Distribution (ITD) Model for Overload Detection and Balancing

The ITD proposed for the load balancing involved a multi-facet approach which begins with the detection of the traffic condition inside the cells. The detected network traffic is controlled with a sector-based traffic control algorithm, and then handover is used to balance the load. Figure 1 presents the lifecycle of the ITD.

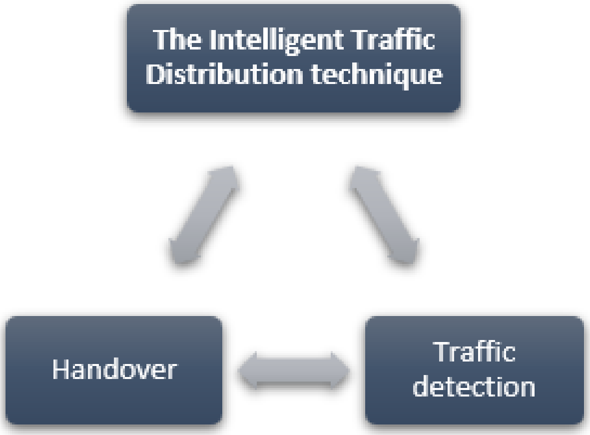


Figure 1: The lifecycle of the Intelligent Traffic Distribution Technique

Figure 1 presents the ITD process flow cycle. First, the network traffic condition was detected using a machine learning algorithm that monitors the network traffic patterns to detect overload. When these signs of overload were detected, a sector-based algorithm inspired by the frequency reuse technique was developed and used to strategically distribute the load from the congested network to other cells, while addressing the issue of interference.

2.21.Traffic detection model

The traffic detection model was developed using methods such as data collection, data processing, artificial neural networks, and training of the neural network, then the generation of a model for the intelligent detection of traffic in the 5G HetNet. Figure 2 presents the process lifecycle. The first step is data collection from MTN Nigeria, considering the HetNet performance for the first 6 months of 2023. The data collection considers the same key performance indicators (KPIs) (Throughput, call drop rate, load factor, Latency, packet loss, etc) applied for characterization and obtains the traffic pattern of the HetNet. The data collected were stored in Comma Separated Value file (CSV) files and then imported as shown in Table 1. The data was processed using an imputation approach to search and replace all missing values and duplicate values in Excel software. The essence is to ensure data completeness and integrity. The total sample size of data collected is 10,190 samples of congestion in the 5G network. The data after imputation was augmented using Gretel application software. The aim is to increase the volume of data collection and help address issues of overfitting during the training of machine learning algorithms. The application software collected the data after importation and then transformed it using 21509 samples of congestion features for 5G HetNet.

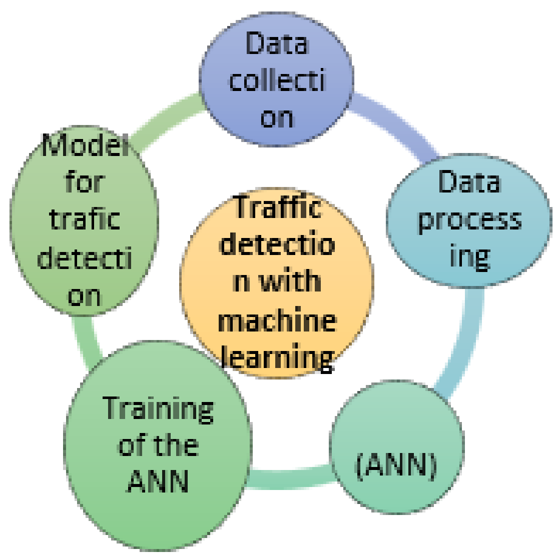


Figure 2: Process lifecycle of the TD

Table 1: First 6 rows of data sample for congestion in 5G network

Time (hh:mm)	Data upload (Mbps)	Load factor	Throughput (Mb/s)	Throughput (%)	Packet loss (%)	Latency (ms)	RSRP (dBm)
06:16.1	820	8.20	488.72	59.6	19.12	228.565	-123.34
09:23.3	830	8.30	197.34	55.8	13.11	163.276	-122.23
12:30.5	840	8.40	124.46	58.9	12.53	122.768	-127.32
15:37.7	890	8.90	511.75	57.5	15.1	364.755	-116.43
18:44.9	850	8.50	209.04	53.6	14.63	134.247	-128.34
21:52.1	870	8.70	145.69	55.7	12.56	115.742	-128.22

2.3 Artificial Neural Network (ANN)

ANN is the machine learning algorithm utilized for the study and trained with the data collected. The ANN is made of neurons interconnected in layers to form a network with the capability to learn and predict network conditions. The input to the neurons (X_n) is channeled through which data enter the network, while the neurons (P_n) have weight and bias. These neurons are interconnected to form the hidden layers which are activated using hyperbolic tangent activation function to produce output Y. Figure 3 shows the neural network, loaded with network data and activation function and ready for training.

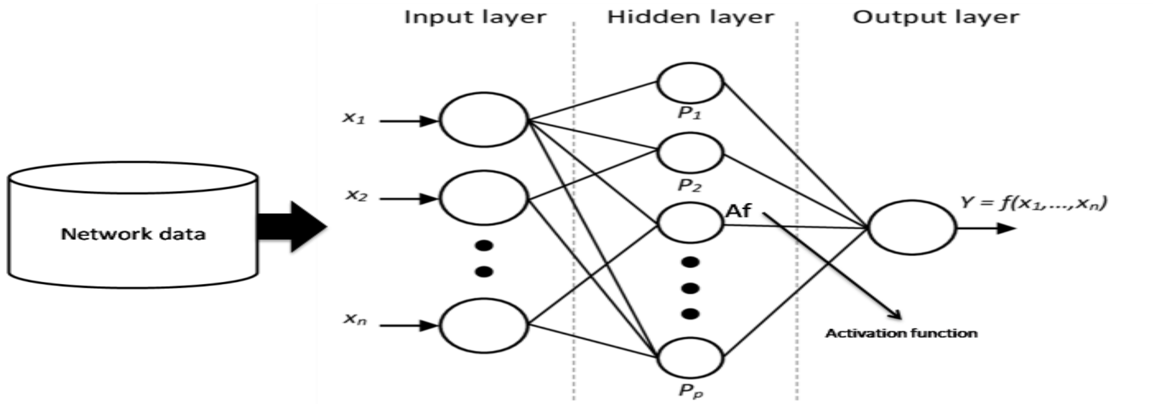


Figure 3: ANN with Network data

The above figure 3 presents the neural network architecture, in designing the neural network the number of dataset attributes informed the number of attributes in the dataset which are 5 inputs while the number of neurons in the hidden layer is 15, and the output is two which represents an overloaded cell or not overloaded. To train the neural network the data was loaded into the neural network as shown in the figure then stochastic gradient descent and regularization techniques were used to train the network.

2.4 Training of the ANN

To train the ANN, the training set was used to initialize the neural network, then the Stochastic Gradient Descent (SGD) in equation 1 adjusts the hyper-parameters of the neural network, which are weight, momentum, bias, and learning rate while monitoring the loss function. This adjustment of hyper-parameters continued iteratively at various epoch (training steps) instances until the gradient loss function becomes consistently close to the targeted value which is $(1e * 10^6)$. During this adjustment process, the regularization model in equation 6 monitors the learning process of the neurons and then assigns a penalty with equation 5 to neurons that are prone to over-fitting and then updates the gradient loss with equation 7. Additionally, while the hyper-parameters are adjusted, test and validation sets were used to evaluate the training performance, considering metrics such as accuracy, loss, precision, and recall respectively. When the best gradient loss is recorded, the training stops at that epoch and then the model for the detection of traffic generated. Figure 4 presents the neural network training block diagram.

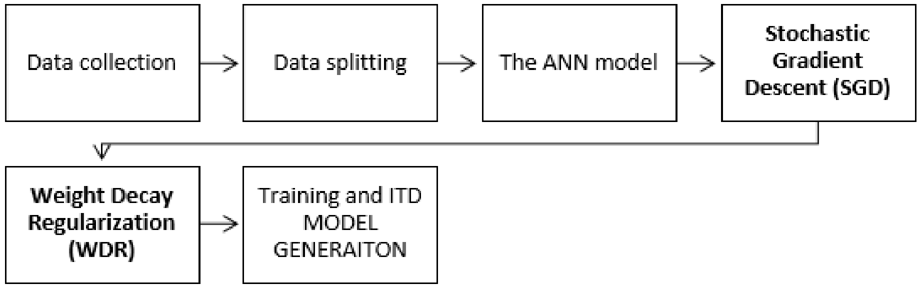


Figure 4: Block diagram of the ANN

The model for the detection of traffic in the 5G network was presented using Figure 5. It identified the network patterns and then compared them with the features of the trained network with traffic. When the features match, then network traffic is detected, else the network tracking process continues until traffic is detected.

2.5 Vertical Handover Process

Handover (HO) was used to balance the load from the congested cell. The process takes three steps which are the process initialization, connection process to the targeted cell, and data transfer process to the target cell (Onuigbo and Ogili, 2023). The process initialization starts by triggering the parameters for the handover request process. Then the population of cells within the HetNet, then the network condition is monitored considering Carrier Signal Information (CSI) such as RSRP, Reference Signal Received Quality (RSRQ) using the rule-based approach (Onuigbo and Ogili, 2023) to determine the most-fit cell for handover and identified as the targeted cell. Then HO request is forwarded to the cell and connections are established with the main cell. Then the sector-based traffic control algorithm was applied to distribute the load from the main cell to the targeted cell. During this process, the proposed Clear Channel Anulling (CAA) was applied to identify channels that are free from adjacent channel interference and congestion. Figure 7 presents the flow chart of the HO process.

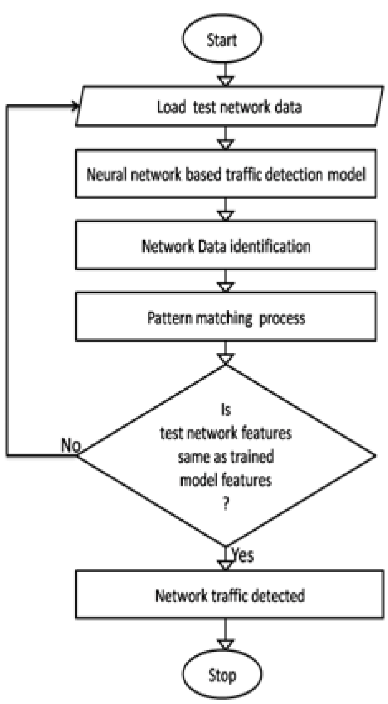


Figure 5 Flow chart of the (TDM)

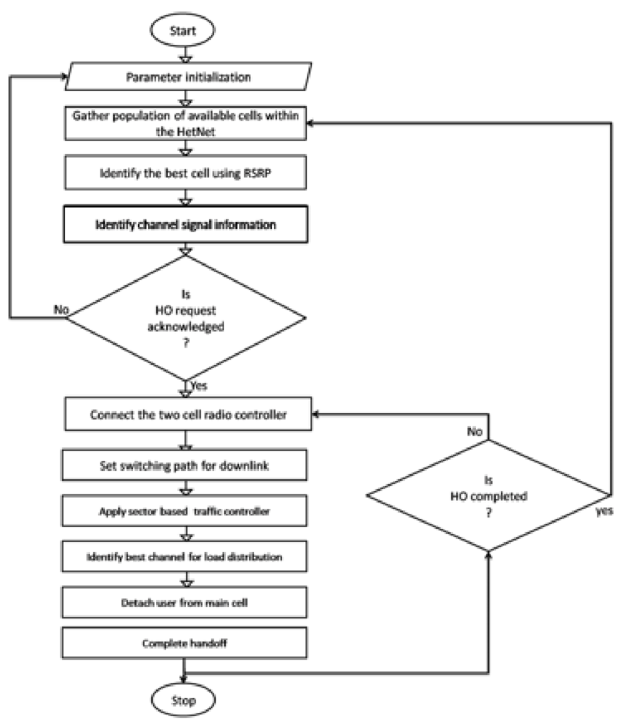


Figure 6: Flow chart of the HO process

In Figure 6, the HO flow chart was presented. The condition of the cell is monitored considering Channel State Information (CSI) such as RSRP, and then the best available cell for the load balancing process is detected and an HO request is sent using the admission control process. Is the selected cell available which is always the case as they are underutilized as identified in the characterization. The targeted cell acknowledges the HO request meaning that it is willing to manage the load, and as a result, the radio network controllers for the two cells are configured and connected automatically for the HO process. During the HO, the sector-based traffic control model was utilized to ensure that the channel used for the HO was free from interference and congestion.

2.8 System Implementation with High-Level Programming

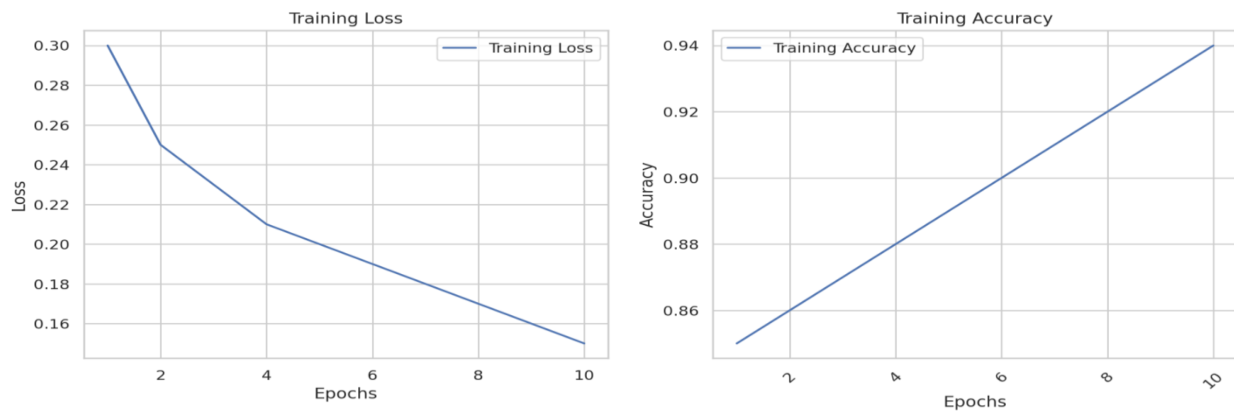
The system was implemented using Python programming language. The TDA was generated with a neural network toolbox. To achieve this, the data collected and processed with Gretel software was imported to the Python environment, and then neural network toolboxes were selected and trained the data using the SGD algorithm. During the training of the ANN, the performance was evaluated and when the best result was obtained, the training stopped automatically and generated the traffic detection model. This model was exported to a Matlab environment and then applied as input to the sector-based algorithm which optimized the handover process during load balancing. The parameters used for the programming are reported in Table 2.

Table 2: Cell Information for the Heterogeneous Network (MTN, Nigeria)

Parameters	**Values**	**The Macro Cell Station**	
Carrier frequency of cell	2.6GHz	**Parameters**	**Values**
System bandwidth	800MHz	Antenna height	20m
Congestion; non-congestion data	161; 839	Transmission power	46dBm (40W)
Shadow fading	9.22dB	Distance in radius for cell range	1-20km
Pathloss exponent	3.79	Traffic load target (Mb/s/km^2)	850
Channel model	3GPP SCM	**Pico Parameters**	**Values**
UE gain UE noise speed	560km/h	Transmission power	22dBm (1W)
Inter-site distance	500m	Reduced transmission power	11dBm(12mW)
Noise spectral density	-175.1dBm/Hz	eNodeB Antenna gain	2dBi
Special sub-frame ratio	2/8 (1 ABS + 1 RPS)	Antenna height	12.8m
Total voice packet used for the simulation	0.6mb	Distance in radius for cell range	47m
Channel model	Typical Urban	Pico Cell Station	
Modulation	64QAM	Parameters	Values
Sub-frame duration	1s	Parameters	Values
Subcarrier number	12	Transmission power	62dBm
Time window size	9	Site type	Single sector
Frequency window size	13	Reduced transmission power	13dBm(14mW)
Specification for VOIP	2 x 88bit	eNodeB Antenna gain	7dBi
Simulation time	16hours	Antenna height	10m

3. Results and Discussion

This section is the results of the neural network training. To train the neural network algorithm, the analyzed data was imported into the algorithm and then split into training, test, and validation sets. The training set was used to train the ANN, applying SGD and regularization. The SDG adjusts the hyper-parameters while the neurons learn the traffic patterns on the network. The regularization technique ensures that over-fitting does not occur during the learning process and ensures that the model generated is generalized. During the training process, accuracy and loss functions were applied to evaluate the model, and the results were reported in Figure 8; while the average accuracy and loss function recorded for the training was reported in Figure 8.



(a) (b)

Figure 7: ANN training performance

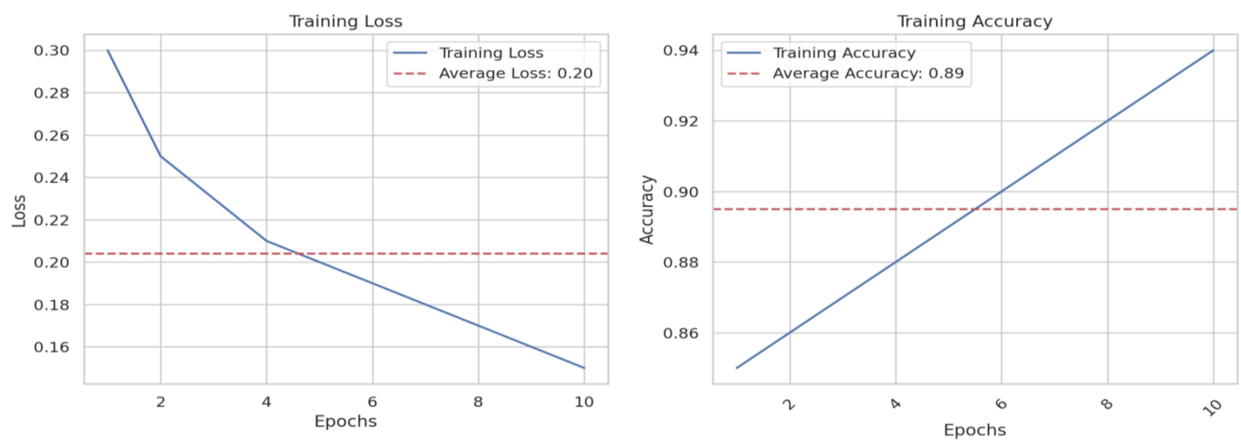


Figure 8: The Average accuracy and loss performance

Figure 7 shows the training performance of the ANN. During the training process, the performance of the gradient loss is evaluated at every epoch, considering the loss function in Figure (a) and also accuracy in Figure (b). This was achieved using the test set to test the neurons and then validate simultaneously until the best version of the TDM was generated. From Figure 8, it was noticed that the evaluation process takes 10 epochs and the best version of the model (which is the point where the least loss and best accuracy was recorded) is at epoch 10. Figure 8 reported the average loss to be 0.20, while the average accuracy reported 0.89. What this means is that during the training and testing process, the accuracy of correctly detecting traffic on the 5G network is 0.89%, while the loss 0.20%. This performance showed that the ANN was able to correctly detect traffic on the 5G network with high accuracy. To validate the result of the training process, a ten-fold cross-validation process was applied as shown in Figure 10.

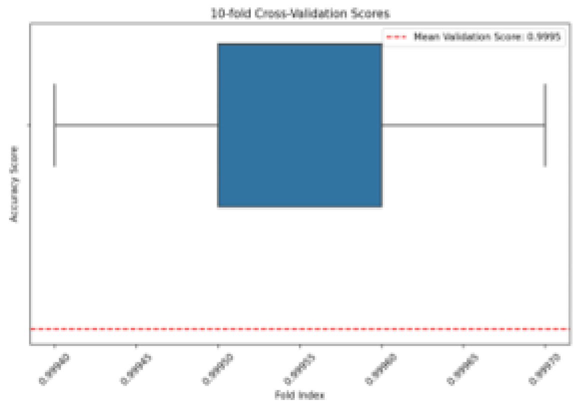


Figure 9: Cross-validation result of the ANN

Figure 10: Average cross-validation

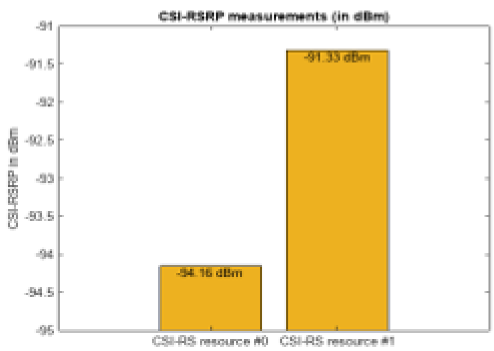


Figure 11: RSRP assessment for HO

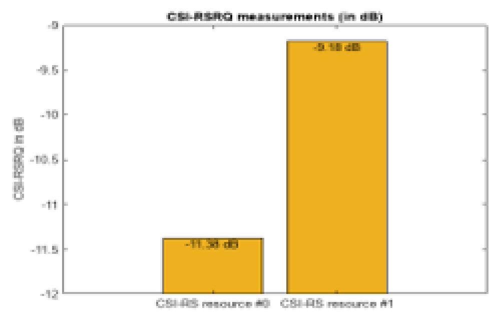


Figure 12: RSRQ assessment for HO process

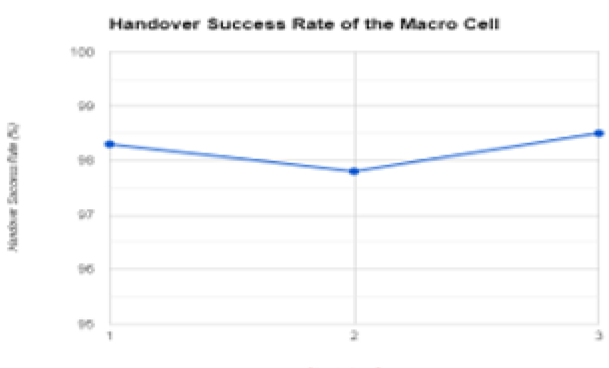


Figure 13: Handover success rate

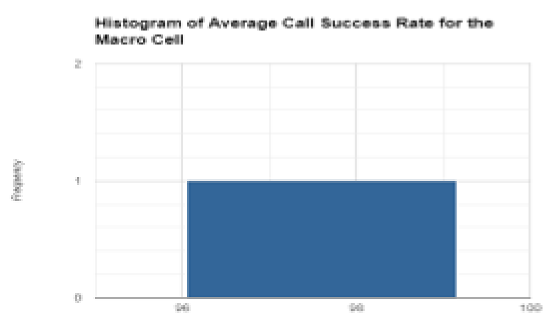


Figure 14: Call success rate

These CSI parameters serve as key indicators, providing valuable insights into the quality of the signals received from neighboring cells. In Figure 11, the RSRP result was presented, offering a depiction of the received power levels from neighboring cells. From the first available cell with the #0 serial number, the RSRP is -94.16 dBm, while for the second available cell within the HetNet with label #1, the RSRP is -91.33 dBm. What this means is that cell 2 is the best for the HO process and was recommended for the load distribution while utilizing the selected channel from the sector-based control model to avoid congestion and interference. Figure 12 showcases the RSRQ assessment for HO. This RSRQ provides information on both the received signal quality and the interference levels. These monitoring steps enable the network to dynamically assess the performance of available cells, paving the way for seamless handover operations based on real-time data and a holistic understanding of the network's condition. The integration of these CSI parameters ensures a robust understanding of the cellular environment, guiding the HO process toward informed decision-making. To support these results, the HO information of the cell was also evaluated considering the handover success rate and call success rate. The analysis was performed using the Nigeria Communication Commission (NCC) standard for HO evaluation in LTE-A, and the results were presented, starting with the handover success rate. The presentation of Figure 13 depicting the HSR of the macro cell at 98.3 % signifies a robust and efficient handover management system within the cellular network infrastructure and underscores the network's capability to transfer ongoing calls across different coverage areas. This high success rate indicates optimal network performance, effective signal handoff, and reliable connectivity for users traversing through the macro cell's coverage zone.

Figure 14 presents the call success rate performance of the cell. The result showed that the average call success rate achieved for the macro cell is 97.6 %. The call success rate serves as a critical metric in evaluating the efficiency and reliability of cellular network services, reflecting the percentage of completed calls compared to the total number of attempted calls within a defined timeframe. A call success rate of 97.6 % for the macro cell indicates a high level of reliability and effectiveness in establishing and maintaining connections for users within its coverage area. This achievement underscores the network's capacity to successfully handle a significant proportion of call attempts, ensuring minimal call dropouts and optimal user satisfaction. The figure provides valuable insights into the performance of the macro cell, highlighting its ability to deliver consistent and dependable communication services to subscribers. It also signifies the effectiveness of the adopted handover algorithm in facilitating seamless and reliable voice communication within the cellular network ecosystem.

Table 3 showcased the performance of the HO with ITD. The average handover success rate is 98.08 %, the average call success rate is 98.39 %, and the average drop call rate is 3.81 %. The handover success rate is consistent across all three cell types, with an average of 98.07 %. The call success rate is highest for macro cells (98.40 %) and lowest for pico cells (98.32 %). The drop call rate is highest for pico cells (1.11 %) and lowest for macro cells (1.07 %). Overall, the call quality is good, with high handover and call success rates and low drop call rates.

Table 4 presented the comparative handover performance with ITD with another sample data collected from MTN Nigeria on a macro cell with an identification number of T4699, coordinate is at Latitude 6.3740’40” N, and Longitude 7.574’90” N, and frequency band for the MTN network is 900MHz to 2600MHz and operated using the concept of frequency reuse.

Table 3: Handover with ITD

Days	**Handover Success (%)**			**Call Success Rate (%)**			**Drop Call Rate (%)**		
Cell types	Macro	Micro	Pico	Macro	Micro	Pico	Macro	Micro	Pico
1	98.077	98.13	98.043	97.05	96.75	96.67	2.4095	1.17	1.12
2	98.082	98.64	98.034	97.12	96.52	96.56	2.4.048	1.06	1.08
3	98.073	98.41	98.045	97.40	96.53	96.65	2.4095	1.09	1.05
Average	98.0773	98.3933	98.0407	97.19	96.6	96.6267	2.4083	1.10667	1.08333

Table 4: Comparative handover performance on macro cell data collected from MTN

Days	HSR with ITD	HSR without ITD	CSR with ITD	CSR without ITD	DCR with ITD	DCR without ITD
1	98.077	91.077	97.05	93.05	2.4095	3.95
2	98.082	92.082	97.12	93.12	2.4.048	3.88
3	98.073	92.073	97.40	92.40	2.4095	3.595
Average	98.0773	91.744	97.19	92.8567	2.4095	3.80833

Table 4 showcased the average values on the network with ITD after three days of testing. The HSR system with ITD demonstrates an average of 98.0773 compared to 91.744 without ITD. This marks a substantial improvement of approximately 6.83 %. Similarly, in the CSR system, the average with ITD reaches 97.19, showcasing a significant enhancement over the average without ITD, which stands at 92.8567. This indicates a percentage improvement of around 4.56%. Moreover, in the DCR system, ITD reported 2.4095 as against without ITD which reported an average of 3.80833, implying that ITD includes a reduction in the DCR OF 36.76 %.

4. Conclusion

This paper has successfully presented an intelligent traffic distribution model for load balancing using an artificial neural network. A mathematical model of the network was presented and data during operation was collected and applied to train neural network algorithms to produce the ITD. The model was integrated with the vertical handover model into the 5G network and evaluated. The results showed that the model was able to detect traffic patterns on the network, and then the handover was initiated to distribute the load and maintain stability on the network. The model was recommended to help improve load balancing in a 5G heterogeneous network.

Reference

Chai R., Dong Y., Ma J., and Chen B. (2011) “An optimal IASA load balancing scheme in heterogeneous wireless networks”. International Journal of Electrical, Electronics and Data Communication, ISSN; 2320-2084; Vol 3; Issue 8. Pp.22-28

Elmosilhy., Noha A., Abd E., Ahmed M., Elmesalawy, Mahmoud M. (2019) "Leveraging Technologies For Growth"; Wireless personal communication 95(4); DOI: 10.1007/s11277-017-4042-5

He X., Ren Z., Shi C., and Fang J. (2016), “A novel load balancing strategy of software-defined cloud/fog networking in the Internet of Vehicles,” vol. 13, no. Supplement2, pp. 140–149.

Kavitha A., Koppala G., and Leela R. (2020) “A Cluster-Based Routing Strategy Using Gravitational Search Algorithm for WSN” Vol. 14, No. 1, pp. 26-39.

Kavitha A., Koppala G., and Leela R. (2020) “A Cluster-Based Routing Strategy Using Gravitational Search Algorithm for WSN” Vol. 14, No. 1, pp. 26-39.

Mohammed S., Pierre C., and Marceau C. (2020) “Reinforcement Learning Algorithm for LoadBalancing in Self-Organizing Networks”. Information Sciences, vol. 179, no. 13, pp. 2232–2248.

Ogili Solomon Nnaedozie, Onuigbo, Chika M., (2023)” Addressing Constraints of Mobility Management in 4g Network Through Hybrid Optimization Technique” IJORTACS; Volume 2, Issue II, February 2023, No. 40, pp. 414-428

Raid S. (2019) “Neural Network Design for Intelligent Mobile Network Optimisation”. Natural Computing, vol. 9, no. 3, pp. 727–745.

Ravyaar J., Ali Y., Pinar K. (2019) “Load balancing in multi-tier cellular 4G using cloud computing”. *IEEE J. Ocean. Eng.* **2019**, *48*, 246–259

Tian, Y.; Zhang, Y.; Zhang, H. Recent Advances in Stochastic Gradient Descent in Deep Learning. *Mathematics* **2023**, *11*, 682. https://doi.org/10.3390/math11030682.

Vasan I. (2019)” This thing called weight decay” https://towardsdatascience.com/this-thing-called-weight-decay-a7cd4bcfccab; accessed 3/12/2024.