Traffic Classification using Deep Learning Approach for End-to-End Slice Management in 5G/B5G

Noor Abdalkarem Mohammedali, Ali Al-Sherbaz, Triantafyllos Kanakis and Michael Opoku Agyeman

Abstract: Network slicing is a key role in future networks. 5G networks are intended to meet different service demands of an application offered to users. 5G architecture is used to match the requirement of the Quality of Service (QoS) by addressing different scenarios in terms of latency, scalability and throughput with different service types. Using machine learning with network slicing allows network operators to create multiple virtual networks or slices on the same physical infrastructure. These slices are independent and customized. Precisely, These slices will be managed dynamically according to the requirements defined between the network operators and the users.

For this research, multi-machine learning algorithms are used to train our model, classify network traffic and predict accurate slice type for each user. After the traffic c lassification, we compared and analysed the performance of various machine learning algorithms in terms of learning percentage, accuracy, precision and F1 score.

Index Terms: 5G, Machine Learning, Network Slicing, services, NFV, SDN, Deep Learning, End-to-End, classification Model

1. Introduction

Future networks are intended to meet different service requirements. For this, 5G systems plan to increase resources offered to users by adding new bands in the spectrum and using new radio access technologies, such as massive multiple inputs multiple outputs (MIMO). 5G networks will not use the monolithic architecture networks, since such architecture is unable to meet different services of the most diverse use cases. Such use cases include, for example, vehicles autonomous, intelligent hospitals and wearable technologies. For that, End-to-end slice isolation was introduced in [1] to isolate the physical infrastructure to virtual networks.

Network virtualization and Softwarization is a trend research area with many open research issues waiting to be addressed. Amongst the main areas that need further investigation are mobility management, resource allocation and slice management [2]. In mobility management, the network manages based on the customer requirement called the Service Level Agreement (SLA) between the service provider and the customer (organization). In addition, there is another agreement between the organization and the end user called Business Level Agreement (BLA). In BLA, the user has limited service and specific criteria. On the other hand, in slice management, the slice (per flow, per service, etc.) is managed dynamically based on the agreement between the customer and the service provider [3].

There are three types of services Network Slicing support eMBB (enhanced Mobile Broadband), URLLC (ultra-Reliable Low Latency Communications) and MIoT (Massive IoT) [12]. So, each type of service is served by a different slice. In eMBB, this service can handle consumer mobile broadband, including a High-Quality Video stream and Fast large file transfers. In URLLC, this service requires high reliability and low latency used in Vehicle-to-Vehicle (V2V), Vehicle to-Infrastructure (V2I), and the device-to-device (D2D) communications concepts [4]. In MIoT, this service handles a large number of IoT devices efficiently and cost-effectively [5]. Within these services, each application has a different priority level: high priority or low priority [6]. In addition, managing the priority of the slice that the user service belongs to and the priority of the service that the end user requires in each slice remains an open issue. Finally, there are two types of slice priority: inter-slice priority and intra-slice priority [7].

To support the most diverse use cases, the 5G networks aim to integrate the concept of Network Slicing in their architecture. Although Network Slicing is a new technology, only recently has it been introduced in wireless networks. Virtualization of the network infrastructure can be performed, for example, from a hypervisor. The concept of Network Slicing in 5G networks is realized thanks to the network software technologies, such as Network Virtualization Functions (NFV) and Software Defined Networks (SDN) [8]. These are intended to bring the benefits of software to networks. The flexibility and modularity are examples of such benefits. In our survey paper [9], we reviewed all these techniques and highlighted the current issues and future directions that could be used to implement network slicing to improve the QoS/QoE in future networks.

Network Slicing can be done in both network layers: Radio Access Networks (RAN) and 5G Core network. Computer resources, spectrum, as well as network functions can be virtualized and distributed to different slices belonging to the 5G networks. Mobile operators have been facing many difficulties to accommodate the network infrastructure traffic of their users. In addition, there was an expressive increase in this traffic in recent years due to the popularization of smartphones and tablets [10].

In Section II, we summarize the background research to develop a deep learning approach for E2E slice management. In Section III, we propose a Random Forest algorithm and classification model for the dataset. In Section IV, we explain and evaluate the performance of the model. Finally, in Section V, we conclude this paper.

2. Related Work

Mobility management in 5G reviewed in [11] to match the demands of the Quality of Service (QoS) for different service types, such as: eMBB, mMTC and URLLC. Their solution tackled different scenarios in terms of throughput, latency and scalability. Inter and intra-slice management and network function placement is mentioned in [12]. The authors highlighted that the network slicing approach needs significant effort when it is used with next-generation mobile networks. In addition, the authors in [13] proposed a framework based on a model of proportional allocation of resources with the objective of realizing the concept of network slicing in multitenant networks. The proposed framework allows for dynamic sharing between the slices, increasing overall tenant performance. Besides that, their framework is generic and is not based on specific cellular technology.

Graph theory is proposed in [14] to manage the inter-slice and save the slice in the queue based on the probability event. Moreover, predictive solutions are provided in their work to evaluate the network and improve the QoS in future networks. A hybrid learning algorithm was proposed in [15] to classify the slices in the 5G system after optimizing the weight function using GS-DHOA, but their work needs further improvement to solve complex problems. Their dataset includes device types, reliability, duration, delay, jitter, bandwidth, speed and modulation type.

In [16], the support vector machine (SVM) algorithm is proposed to select slice features according to the IoT services. An unsupervised algorithm is used for grouping similar applications in one cluster called sub-slice clustering. On the other hand, K-means has some limitations in terms of latency. On the other hand, in [17], K-means is used for clustering three slice types after identifying 22 features from their dataset. Furthermore, with their classification results, they achieved high accuracy for all selected algorithms. In [18], Machine learning model proposed for throughput prediction.
They predicted the throughput for non-standard 5G networks, and their accuracy achieved 84% and 93%. On the other hand, Chi-square method is used for nonlinear dataset and their accuracy result was 99% for 25 features as explained in [19].

In [20], the author proposed a survey on how we can manage the slice using machine learning. Reinforcement learning and Neural networks algorithms are summarized in their work. Transfer learning is proposed in [21] to accelerate the 5G resource allocation using a deep reinforcement learning algorithm on the radio access network. On the other hand, deep reinforcement learning is implemented [22] to examine the effectiveness of the slice resources in 5G networks in terms of utilization and delay when considering the relationship between the node and the impact of surrounding nodes' resources.

In this paper, we will use a deep learning algorithm and convolutional neural network (ConvNet) for traffic classification. In our model, we have eight features as input, four-layer as a hidden layer and the 5G slice types as output; we will discuss more it in the next section. This research is a continuous work for our survey paper in [9]. We reviewed all the state-of-art techniques that could be used to control the slices for the 5G systems.

3. The Proposed Model

In this section, we will use a filtered structured and cleared dataset which is used for training and testing the model. The original file for the dataset did not contain a header for each column, but there is an excel file that explained the dataset accurately. In [23], the authors explained how they prepared and collect the dataset to fit the 5G networks. In our model, after using the dataset, operating the model and separating the training model into many phases, the accuracy for the prediction of slices increased and the number of losses decreased. Furthermore, the number of features in the dataset was eight and we added a new column for the location of the user in the 5G network based on the traffic we generate to check the performance of the network when the user is within the home network and the visited network.

In this paper, we trained the 5G model using a Deep Learning algorithm to train and predict slice types for a device based on the information that calculated from previous connections. the dataset is treated with a high-level API called (Keras) to build and train them in TensorFlow. Moreover, this machine learning model would be based on supervised learning because the dataset was big and structured. As a

result, the Random Forest algorithm and ConvNet are used for the traffic classification. Furthermore, various parameters are utilized to determine the network slicing: slice type, bandwidth, throughput, latency, equipment type, mobility, reliability, isolation and power.

The goals for this model: A- Select a slice for a device. B- Select enough resources for the slice based on the traffic prediction. The dataset features description:

1) Device types: This column contains a group of devices: Mobile, VR, Healthcare, IoT, Gaming and Industry 4.0.
2) Device Category: 5G and LTE (4, 5, up to 20). 3) Technology Supported in terms of LTE, IoT, LTE-M, NB-IoT and 5G.
4) Duration: Connection duration in day and time.
5) Guaranteed Bit Rate (GBR) and Non-GBR.
6) Packet Loss Rate: Reliability for sending and receiving packets. For example: 0.000001 and 0.01.
7) Packet delay budget: Latency. For example: 10ms, 50ms and 75ms.
8) Roaming: Home and visited slice network.
9) Slice Type: eMBB, URLLC, mIoT and V2V.

4. Result and Decussion

From the dataset features, we notice that it is a collection of Heterogeneous wireless networks (HetNets). In this case, some challenges appear with HetNets as explained in [24]. First, we need to know the HetNets patterns to obtain an accurate mathematical model to enhance the performance of the HetNets and network slicing. Second, how can we meet the QoS demands of the Slice based on SLA? Finally, how can we assign the spectrum dynamically to ensure the SLA for the slices?

The dataset held 66K rows and 9 columns. The rows stood for the 5G slice parameters, while the 8 columns held the Key Performance Indicators (KPI), and the last column corresponded to the slice types of the 5G networks. Each feature in the dataset had a label. The KPI parameters symbolized input and the last column symbolized the 5G slice types as output.

For the traffic classification, we would use popular machine learning algorithms to find the best algorithm that fits the dataset, such as:

1) Na¨ıve Bayes: The classification in this algorithm uses Bayes' theorem to count the probability of data that belongs to a particular type.
2) Support Vector Machine(SVM): It works with linear and non-linear classifiers. SVM makes predictions based on the support vectors.
3) Neural network (NN): It works with a non-linear algorithm.
4) Gradient Boosted Tree (GBT): this algorithm trains multiple trees to reduce the cost function.
5) Random Forest (RF): It is choose a number of trees to do the classification. This algorithm will predict the final class based on the majority votes.

In the training stage, all the data (X _train) is sent to the machine learning algorithm to learn and come up with the correct answer for (y train). The algorithm uses the following formula for that: f(X train) = y train. For the prediction, the algorithm took the output (y train) and applied it to another formula y _train = f(X train). In this case, the system will be able to predict the 5G slices for any new input that contains all the 5G slice parameters. In the validation stage, the crossvalidation technique is used to evaluate the model and check if it is working dynamically by choosing the correct slice type when we add new data. Afterwards, the model evaluation will be done by adding X test to check if the model could predict the correct y test.

Accuracy is used to view the relationship between the number of correct predictions and the total number of predictions. The accuracy formula is given below to evaluate the performance of the predictive model. Where, the accuracy formula contains True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) [25].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

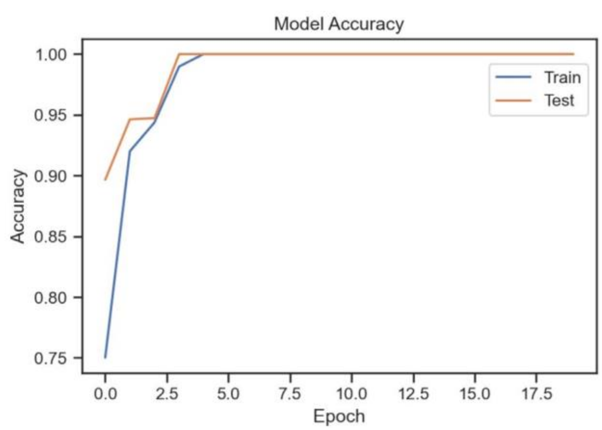

Fig. 1. Accuracy

Figure 1 shows the number of the slice prediction increased and in Figure shows 2 the number of losses decreased.

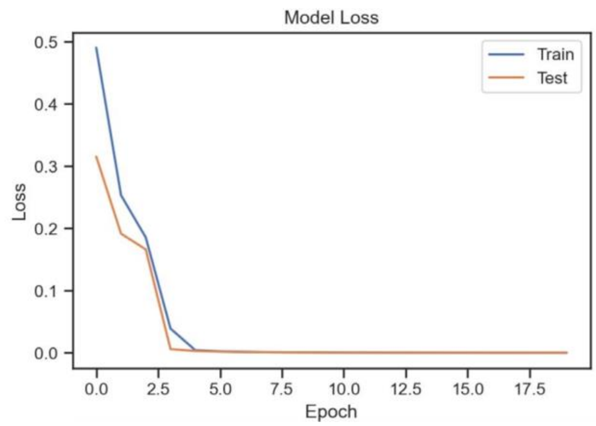

Fig. 2. Loss

Precision: The number of the TP, when it predicts yes, how often is it correct? [25]. The precision formula is given by:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall or sensitivity of the model when it detected positive values. The Recall formula is given by [25]:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

	precision	recall	f1-score	support
H_Net	1.00	1.00	1.00	141
V_Net	1.00	1.00	1.00	127
accuracy			1.00	268
macro avg	1.00	1.00	1.00	268
weighted avg	1.00	1.00	1.00	268

Fig. 3. Measures the precision, recall and F1

F1 score: This score represents the average of precision and recall. The F1 formula is given by [25]:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \quad (4)$$

After we applied formulas on the dataset, the result shown in the Figure 3

Random Forest was implemented using Python with 10 trees. Small tree shown in Figure 4.

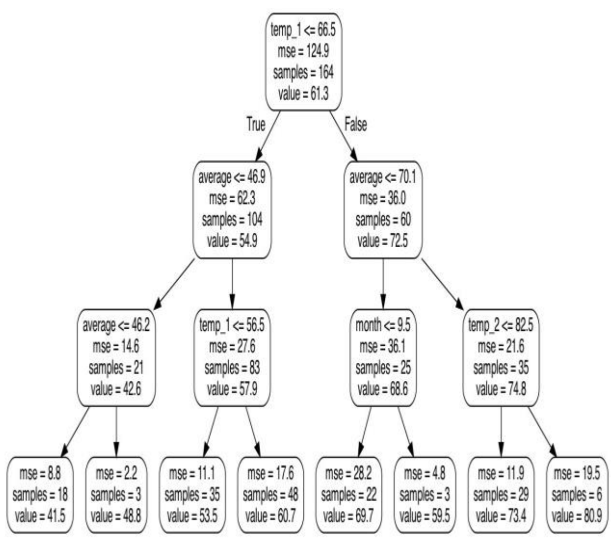

Fig. 4. Small Tree from the Decision Tree of the Random Forest

A confusion matrix is a relationship between the predicted values and the actual values. In this section, a confusion matrix is used to check the performance of the classified model after the prediction.

The matrices from the RF algorithm as shown in Figure 5 and 6. Figure 5 shows the confusion matrix for the user location in a home network and in the visited network while Figure 6 shows the confusion matrix for the 5G slice types.

The confusion matrix for the classification model. The number of correct predictions shown and for the models that are not fit with our model, the number of wrong predictions appears in the matrices. Confusion matrix terms:

1) True Positives (TP): we predicted yes; they have 5G slice type.
2) True Negatives (TN): we predicted no; they have 5G slice type.

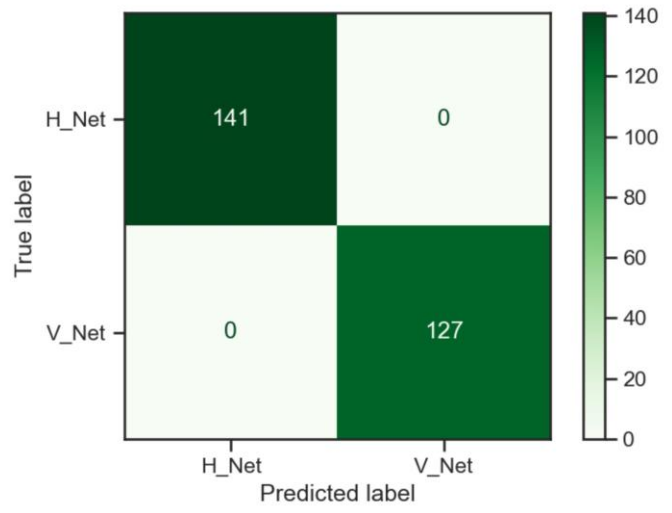

Fig. 5. Confusion matrix for home and visited network

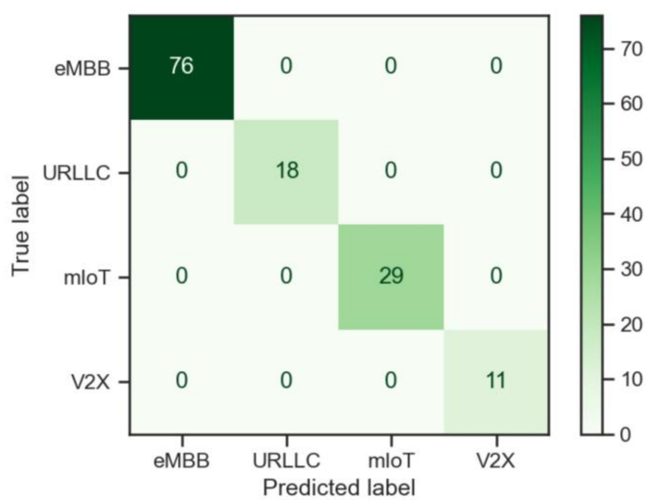

Fig. 6. Confusion matrix for the slice type

3) False Positives (FP): we predicted yes; but they don't have a 5G slice type.
4) False Negatives (FN): We predicted no, but they do have a 5G slice type.

All the dataset features were trained with Trees, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Ensemble, Neural Network and Kernel as is shown in Table I. Each model had accuracy, prediction's speed and time for the training. In terms of accuracy, the majority of classification model had 100% except for boosted trees, kernel, subspace discriminant, and RUSBoosted trees.

5G services will be classified and predicted using supervised machine learning algorithms as is shown in Table I. Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Kernel. MATLAB was used for the compared and contracted. K-Folds was applied using K = 5 for the cross- validation technique.

TABLE I CLASSIFY THE ML MODEL TO FIT THE PROTOTYPE

Class	**Classification Model**	**Accuracy**	**Prediction Speed (obs/sec)**	**Train Time (sec)**
Trees	Fine	100	400000	1202.9
	Medium	100	1300000	980.04
	Coarse	**100**	**740000**	**2.5542**
SVM	Linear	100	410000	27.097
	Quadratic	100	240000	52.742
	Cubic	100	230000	81.596
	Fine Gaussian	100	5600	680.26
	Medium Gaussian	100	150000	709.69
	Coarse Gaussian	100	160000	739.51
KNN	Fine	100	17000	757.04
	Medium	100	9600	785.85
	Coarse	100	3300	868.16
	Cosine	100	2800	972.43
	Cubic	100	10000	999.02
	Weighted	100	9800	1027.2
Ensemble	Boosted Trees	50.3	930000	1029
	Bagged Trees	100	160000	1042.1
	Subspace Discriminant	85.6	61000	1053.2
	Subspace KNN	98.8	920	1341.2
	RUSBoosted Trees	50.3	810000	1342
Neural Network	Narrow	100	650000	1347.5
	Medium	100	680000	1353.3
	Wide	100	480000	1367.3
	Bilayered	100	680000	1374.7
	Trilayered	100	600000	1385.2
Kernel	SVM	94.7	22000	1544
	Logistic Regression	92.7	22000	1616.5

For the classification, MATLAB is used to check the best algorithms for the dataset. Table I shows different classification models. With compare and contrast, we can choose the best model for our data to do the classification. In addition, the Trees, SVM, KNN and neural network had best accuracy which is 100%, while Subspace Discriminant had 85% and the less accuracy was Boosted and RUSBoosted Trees as shown in the Table I. Furthermore, choosing the best model to work with network slicing will depend on the accuracy and model training time as is shown in Table I Coarse model has less time for the training competing with other models. In this case, Coarse is the best model and will be applied for future traffic to predict the 5G services with high accuracy and less training time also the computational power will be reduced to decrease the energy consumption.

We had four slice types need to be classified based on the selected features. The confusion matrices for the algorithms as is shown in Figure 7, 8 and 9.

Kernel was used as a numeric predictors for Na¨ıve Bayes. In addition, the training time for this model was 1834.5 sec. After the training, the accuracy for the validation was 94.5%, the prediction speed was 150 obs/sec and the total cost was 3696. The confusion matrix for the slice types prediction using Na¨ıve Bayes as is shown in Figure 7.

The accuracy for the Ensemble model was 85.6 % and the training time for this model was 1053.2 sec with total cost was 9652. Furthermore, learner number was 30 and the confusion

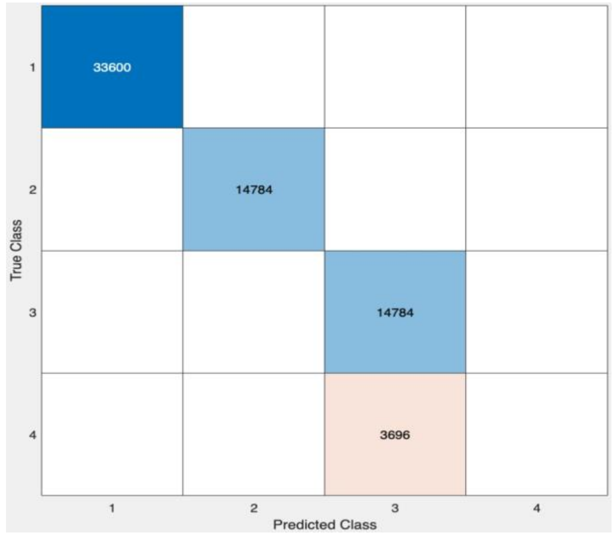

Fig. 7. Na¨ıve Bayes

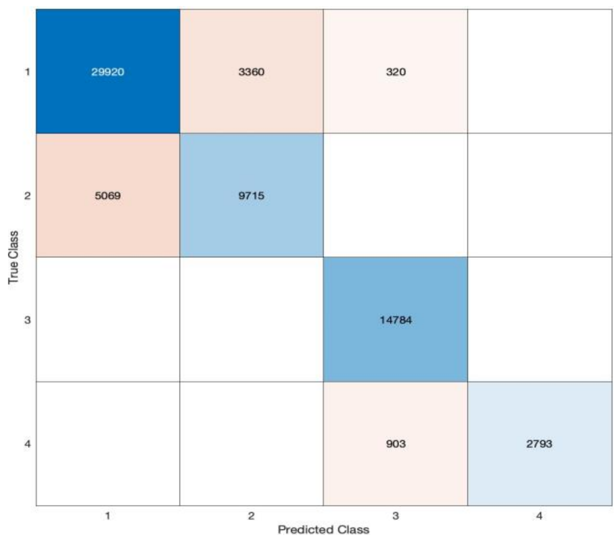

Fig. 8. Ensemble

matrix for the slice types using Subspace Discriminant as is shown in Figure 8.

After the SVM kernal model trained, the total cost was 3534 and the confusion matrix for the slice types using SVM Kernel as is shown in Figure 9.

5. Conclusions and Future Work

In this paper, we compared different algorithms for our model with the selected features. A public dataset was used in this research rather than a self-collected dataset which helped the model to enhance the performance of the classification by knowing the previous data and adopting new information during the learning phase. The dataset held 66K rows and 9 columns as discussed before to choose a good algorithm to fit the model. Each feature in the dataset had a label. The KPI parameters symbolized input and the last column symbolized

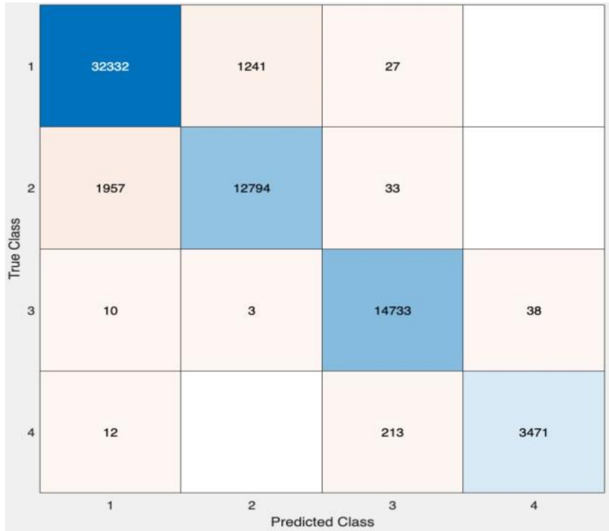

Fig. 9. SVM Kernel

the 5G slice types as output. The accuracy in the Coarse model was 100% and the training time was 2.5542 sec which indicates that this model will give a good result for the realtime traffic classification. For future work, we will deal with real-time traffic and create a robust machine learning model to deal with our dataset and choose the slices dynamically depending on the user requirements. The real-time data will be collected from the End-to-End 5G network in our lab. Further, the computational power will be considered in the training and prediction stage to reduce energy consumption. Finally, A deep reinforcement learning algorithm will be used for slice prediction and classification to improve the accuracy and recall rates in future networks.

Acknowledgement

This research was funded by the Ministry of Higher Education and Scientific Research in the Republic of Iraq and Centre for Smart and Advanced Technologies (CAST) at the University of Northampton to sponsor Noor A. Mohammedali to pursue her PhD research.

References

[1] Z. Kotulski, T. W. Nowak, M. Sepczuk, and M. A. Tunia, “5g networks: Types of isolation and their parameters in ran and cn slices,” *Computer Networks*, vol. 171, p. 107135, 2020.

[2] S. Cherrared, S. Imadali, E. Fabre, G. Gossler, and I. G. B. Yahia,¨ “A survey of fault management in network virtualization environments: Challenges and solutions,” *IEEE Transactions on Network and Service Management*, vol. 16, no. 4, pp. 1537–1551, 2019.

[3] A. Papageorgiou, A. Fernandez-Fern´ andez, L. Ochoa-Aday, M. S.´ Pelaez, and M. S. Siddiqui, “Sla management procedures in 5g slicing-´ based systems,” in *2020 European Conference on Networks and Communications (EuCNC)*. IEEE, 2020, pp. 7–11.

[4] Y.-B. Lin, C.-C. Tseng, and M.-H. Wang, “Effects of transport network slicing on 5g applications,” *Future Internet*, vol. 13, no. 3, p. 69, 2021.

[5] S. Wijethilaka and M. Liyanage, “Survey on network slicing for internet of things realization in 5g networks,” *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 957–994, 2021.

[6] S. A. AlQahtani and A. S. Altamrah, “Supporting qos requirements provisions on 5g network slices using an efficient priority-based polling technique,” *Wireless Networks*, vol. 25, no. 7, pp. 3825–3838, 2019.

[7] A. Othman and N. A. Nayan, “Efficient admission control and resource allocation mechanisms for public safety communications over 5g network slice,” *Telecommunication Systems*, vol. 72, no. 4, pp. 595–607, 2019.

[8] H. Zhang, N. Liu, X. Chu, K. Long, A.-H. Aghvami, and V. C. Leung, “Network slicing based 5g and future mobile networks: mobility, resource management, and challenges,” *IEEE communications magazine*, vol. 55, no. 8, pp. 138–145, 2017.

[9] N. A. Mohammedali, T. Kanakis, M. O. Agyeman, and A. Al-Sherbaz, “A survey of mobility management as a service in real-time inter/intra slice control,” *IEEE Access*, vol. 9, pp. 62533–62552, 2021.

[10] M. Peng, Y. Li, Z. Zhao, and C. Wang, “System architecture and key technologies for 5g heterogeneous cloud radio access networks,” *IEEE network*, vol. 29, no. 2, pp. 6–14, 2015.

[11] N. Akkari and N. Dimitriou, “Mobility management solutions for 5g networks: Architecture and services,” *Computer Networks*, vol. 169, p. 107082, 2020.

[12] A. A. Barakabitze, A. Ahmad, R. Mijumbi, and A. Hines, “5g network slicing using sdn and nfv: A survey of taxonomy, architectures and future challenges,” *Computer Networks*, vol. 167, p. 106984, 2020.

[13] P. Caballero, A. Banchs, G. De Veciana, and X. Costa-Perez, “Network´ slicing games: Enabling customization in multi-tenant networks,” in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.

[14] B. Bordel, R. Alcarria, D. Sanchez-de Rivera, and´ A. S´ anchez, “An inter-´ slice management solution for future virtualization-based 5g systems,” in *International Conference on Advanced Information Networking and Applications*. Springer, 2019, pp. 1059–1070.

[15] M. H. Abidi, H. Alkhalefah, K. Moiduddin, M. Alazab, M. K. Mohammed, W. Ameen, and T. R. Gadekallu, “Optimal 5g network slicing using machine learning and deep learning concepts,” *Computer Standards & Interfaces*, vol. 76, p. 103518, 2021.

[16] S. K. Singh, M. M. Salim, J. Cha, Y. Pan, and J. H. Park, “Machine learning-based network sub-slicing framework in a sustainable 5g environment,” *Sustainability*, vol. 12, no. 15, p. 6250, 2020.

[17] L.-V. Le, B.-S. P. Lin, L.-P. Tung, and D. Sinh, “Sdn/nfv, machine learning, and big data driven network slicing for 5g,” in *2018 IEEE 5G World Forum (5GWF)*. IEEE, 2018, pp. 20–25.

[18] D. Minovski, N. Ogren, C. Ahlund, and K. Mitra, “Throughput prediction using machine learning in lte and 5g networks,” *IEEE Transactions on Mobile Computing*, 2021.

[19] F. Xie, D. Wei, and Z. Wang, “Traffic analysis for 5g network slice based on machine learning,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2021, no. 1, pp. 1–15, 2021.

[20] B. Han and H. D. Schotten, “Machine learning for network slicing resource management: a comprehensive survey,” *arXiv preprint arXiv:2001.07974*, 2020.

[21] A. M. Nagib, H. Abou-Zeid, and H. S. Hassanein, “Transfer learningbased accelerated deep reinforcement learning for 5g ran slicing,” in *2021 IEEE 46th Conference on Local Computer Networks (LCN)*. IEEE, 2021, pp. 249–256.

[22] N. He, S. Yang, F. Li, and X. Chen, “Intimacy-based resource allocation for network slicing in 5g via deep reinforcement learning,” *IEEE Network*, vol. 35, no. 6, pp. 111–118, 2021.

[23] A. Thantharate, R. Paropkari, V. Walunj, and C. Beard, “Deepslice: A deep learning approach towards an efficient and reliable network slicing in 5g networks,” in *2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE, 2019, pp. 0762–0767.

[24] H. Zhang, S. Xu, S. Zhang, and Z. Jiang, “Slicing framework for service level agreement guarantee in heterogeneous networks—a deep reinforcement learning approach,” *IEEE Wireless Communications Letters*, vol. 11, no. 1, pp. 193–197, 2021.

[25] C. A. U. Hassan, M. S. Khan, and M. A. Shah, “Comparison of machine learning algorithms in data classification,” in *2018 24th International Conference on Automation and Computing (ICAC)*. IEEE, 2018, pp. 1–6.