DeepLevel : Mobile Traffic Levels Prediction of Urban Area Using Deep Learning

Ocan Aydın

August 15, 2023

Abstract

As mobile communication continues to growing, there is a need to improve QoS with respect to Service Level Aggrements (SLA). However, this operation is managed by humans and it is not the best way from effectiveness perspective. At this point, DeepLevel focuses to predict mobile traffic levels of urban area using deep learning and it aims to shed light to new mobile technologies especially 5G network slicing. Our work includes exploratory data analysis (EDA) and deep learning for prediction of traffic levels. Our model reached an accuracy rate which is higher than %99.

Keywords : Mobile network, 5G network slicing, deep learning, intelligent systems.

1 Introduction

Mobile communication is everywhere in today's world, actually it became one of our body parts. Therefore, mobile networks continue to evolving from 4G to 5G and 5G will give us higher speed, lower latency and greater capacity. It has three main innovations : Enhanced Mobile Broadband (eMBB), Ultra Reliable and Low Latency Communications (URLLC), Massive Machine-Type Communications (mMTC). eMBB is a concept that focuses on the speed,capacity and mobility to realise use cases such as hd video streaming, augmented reality (AR) and virtual reality (VR). Second concept is URLLC supports use cases which require high network reliability, low latency. Remote surgery, autonomous driving can be given as an example of URLLC needed cases. The third key concept is mMTC will helps us to connect millions of IoT devices to network with its enormous broadcasting area. To use it effectively these services, we need more accurate and robust mobile traffic level prediction ways.

2 Exploratory data analysis (EDA)

The dataset using in this paper is taken from [Ita15]. It includes features such as incoming-outgoing sms and calls. Also it has an internet activity feature and total activity feature which is the sum of other features. Our target variable is traffic level which is assigned to 1-6 with respect to total activity increasing by 20.

2.1 Traffic level variations by each day

The figure 1 shows the traffic level variation for each day in a week. We see that traffic level variation in weekdays are almost same and level 6 is most common level and level 1 follows it. We can think that level 6 is most used level in weekdays due to working time and level 1 is the second one due to old people and housewifes who do not go to work. In weekend, the traffic level patterns is changing and level 1 takes the first place.This situation can be explained there is no working hours in weekends.

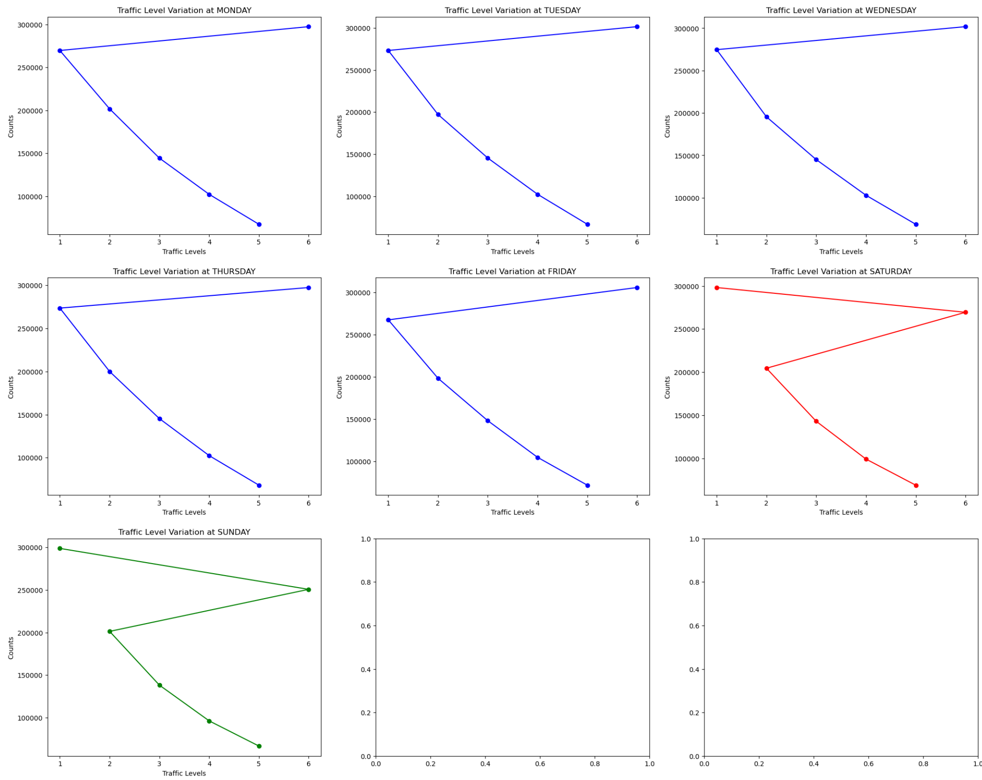


Figure 1: Traffic level variation by each day.

2.2 Total activities by each cell

The figure 2 shows most 5 crowded cells total activity as a sum of one weekend and this is an important feature to extract meaningful information about urban area.

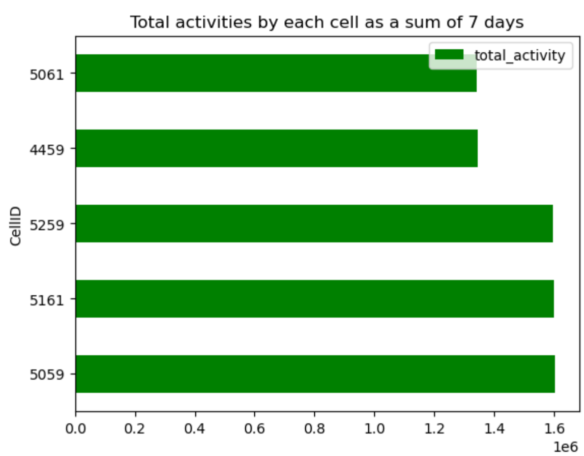


Figure 2: Total activities by each cell ID as a sum of one week.

2.3 Milan urban area and top attractions

We have an Milan city 100x100 grid area which is published in [Ita15]. The figure 3 is created using geojson library and it shows the 100 x 100 grid area of Milan city.

Figure 3: Milan city 100x100 grid area

Using most crowded cells which is given in figure 2, we can find these cells places in Milan city with the help of figure 3. Then we can find the most 5 attractions in Milan city by using cells total activities as shown in figure 4.

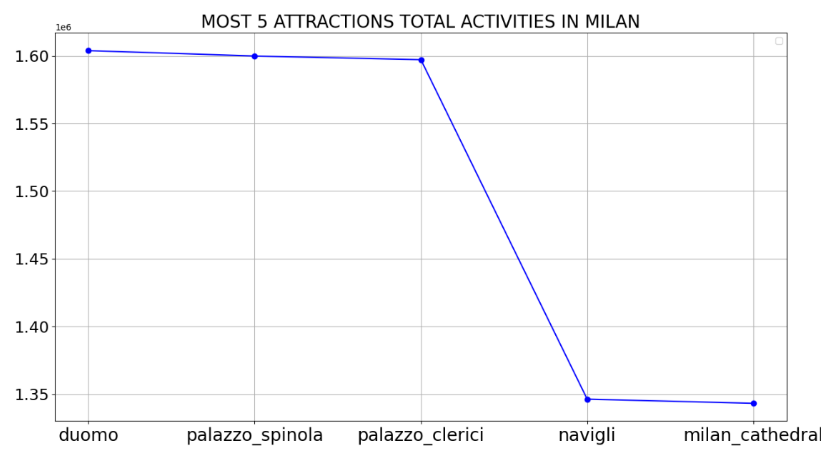


Figure 4: Top 5 attractions in Milan city

2.4 Sms-call activities in top attractions

To understand user behaviours in top attractions, we can look incoming-outgoing sms and call activities. We can also use internet activity in here,but we do not do this because internet activity volume much bigger than sms and call activities so that comparison between these features is not realistic. Figure 5 shows sms and call activities for each attraction in each day. We see that sms-call activities are going down in weekends. After this result, we can say that people make calls or send sms for working issues.

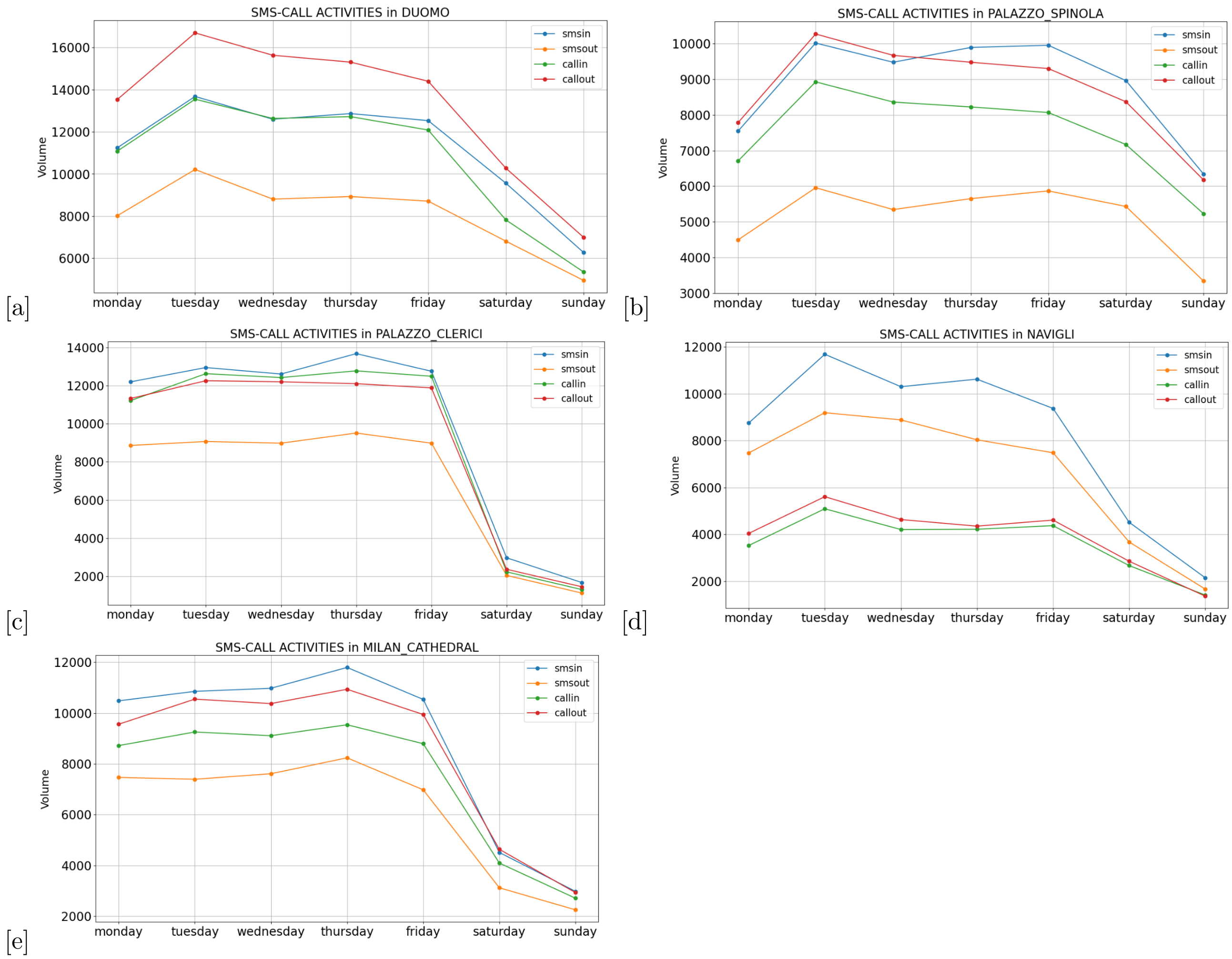


Figure 5: Sms-call activities in top attractions for every day.

2.5 Total activity by hours for each day

Figure 6 shows the total activity by hours for each day in a week. In figure 6, weekdays curves are almost same with each other.This situation can show us that people's usage of mobile connection do not change significantly between weekdays. In weekend, curve is changing and there is a different usage of mobile connection behaviour from weekdays. First of all, total activity is increasing in saturday and sunday evenings and nights. Maybe people watching movies in weekend nights and this can increase total activity. Many outcome can be extracted from this situation.

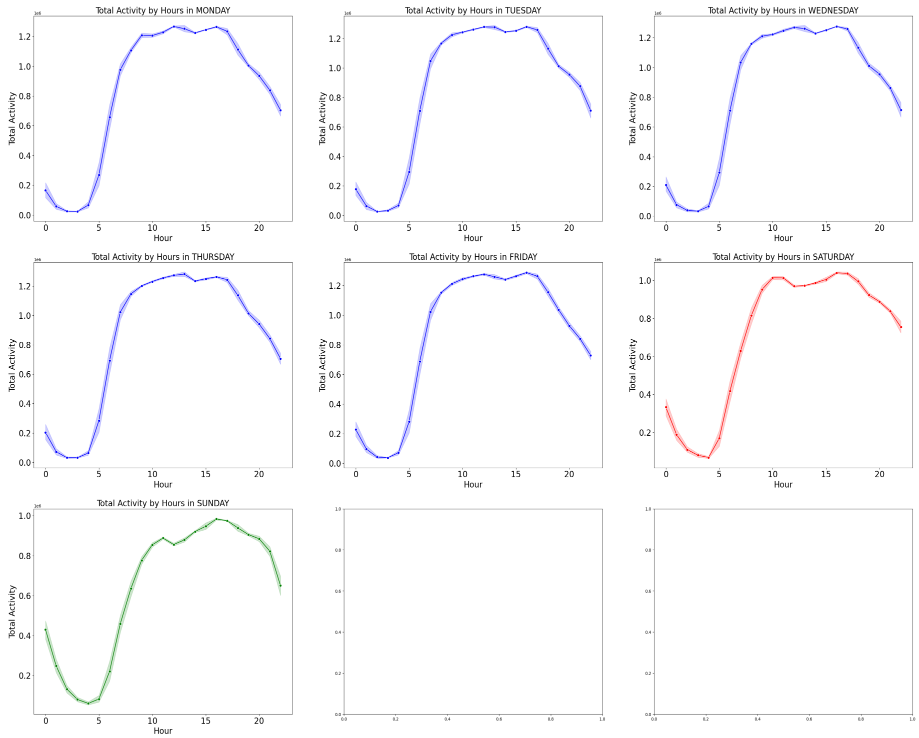
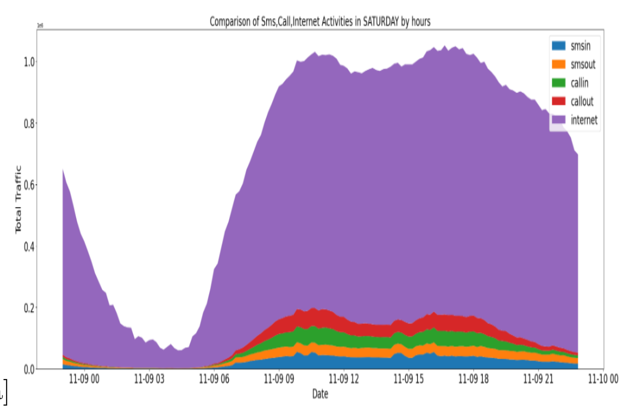


Figure 6: Total activity by hours for each day.

2.6 Volume of activities in weekend

Figure 7 shows the total volume of sms-call-internet activities at saturday and sunday. It shows the activities in whole day with hours. Here, we can clearly see that volume of internet is the biggest one with respect to others. Also, volume of all activities are decreasing in 3 AM and this is too normal due to sleeping hours. From an different perspective, we see that calling activities have more volume than sms activities. Maybe, it happens because of sms usage is high in working times.

[a]


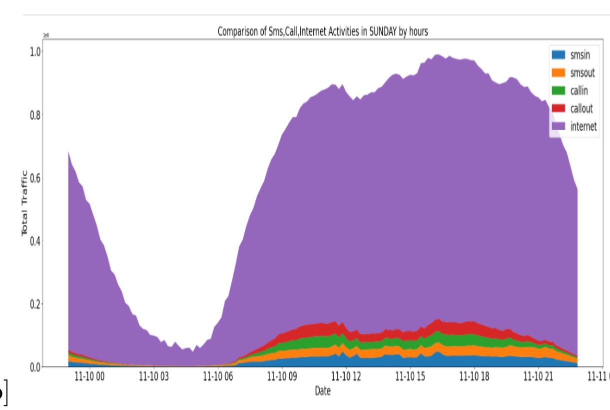
[b]


Figure 7: Total volume of sms-call activities in weekend.

2.7 Total activity and internet activity in country codes

Figure 8 shows top 5 country by total activity and internet activity. Switzerland is the top one because it is so close to Milan city and we can extract that some italian citizens prefer to live in Switzerland or their relatives are living there. The second one is France,it is also meaningful because they have border and so close to Milan. As known already, China and Russia too far from Italy but they are 3 and 4 in the list. Reason for that, they have really big populations and their industry is developed, so that italian citizens might prefer these two places to study and live.

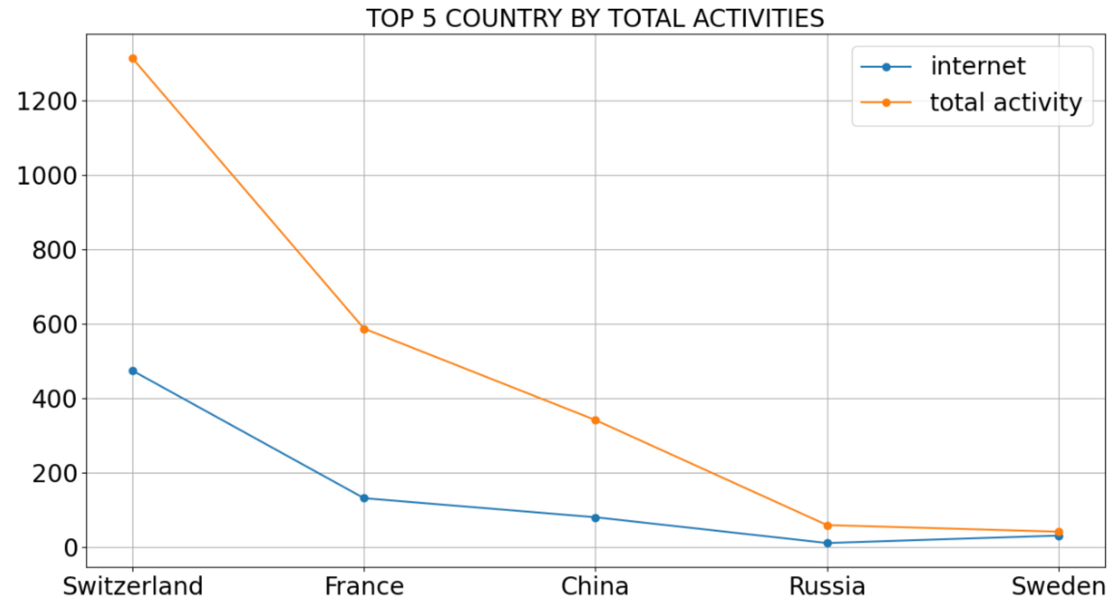


Figure 8: Top 5 country by total activities.

3 Feature extraction and data preprocessing

In this section, feature extraction and data preprocessing part will be explained.

3.1 Feature extraction

For feature extraction, we used pearson correlation matrix to select which features we are going to use and we removed datetime and country codes columns from data to get more robust dataset.

3.2 Data preprocessing

In data preprocessing part, we used a standardization process of features to get more scaled features. Using standardization, we removed the mean and scaling to unit variance. In addition to this, we split dataset as train,test and validation. Train dataset is %60, test dataset %20, validation dataset has %20 of whole dataset.

4 DeepLevel architecture

To predict the mobile traffic levels of Milan urban area, we propesed DeepLevel deep learning architecture. It includes one input layer with 8 neurons and relu activation function, two consecutive hidden layers which have 4 and 3 neurons with relu and tanh activation function and the output layer has 7 neurons with softmax activation function for multi class classification. In compile settings, adam optimizer is used with 0.01 learning rate. The architecture is shown below in Figure 9.

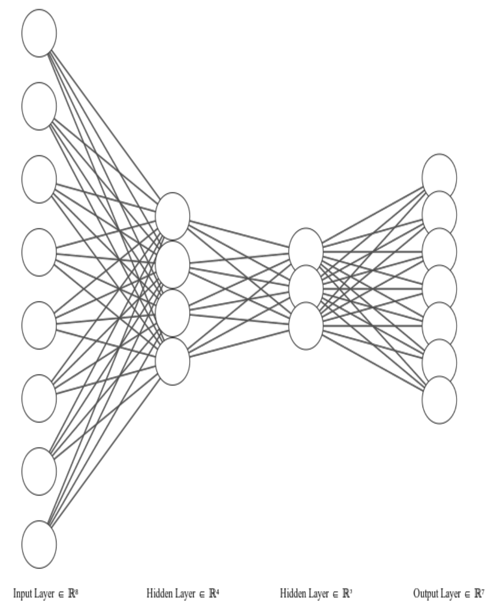


Figure 9: DeepLevel architecture.

4.1 Training

Training part of model using DeepLevel is completed with two different ways. The first one is training same but individual models with each day separately (monday,tuesday,wednesday etc.). The second one is training one model with all days in once (weekly).

4.1.1 Same but individual models for each day

In the first approach, we trained same but individual models for each day . It means that same network architecture is used in each day but they do not affect each other. However , every day trained with 10 epochs and 256 batch size with early stopping. It tooks 11 minutes 14 seconds for total and 1 minutes 36 seconds approximately for each day. Training and validation losses are shown in figure 10. In the figure 10, all losses are going to down and model are successful for each day. But in specific, saturday and sunday models are the best ones hence we can say that weekend days have more identifiable mobility patterns.

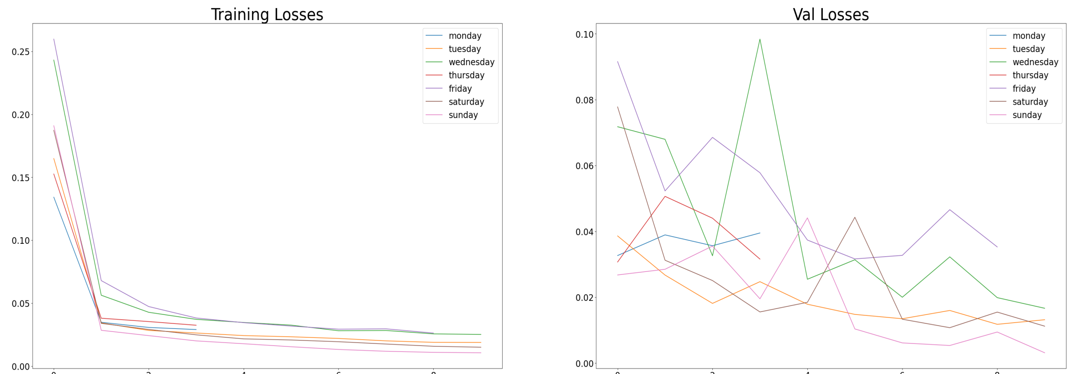


Figure 10: Training and val losses for each day model.

Figure 11 shows the training and validation accuracy for each day models. In accuracy part, we can see the same results as loss part. Weekend days models have more robust weights and more identifiable mobility patterns.

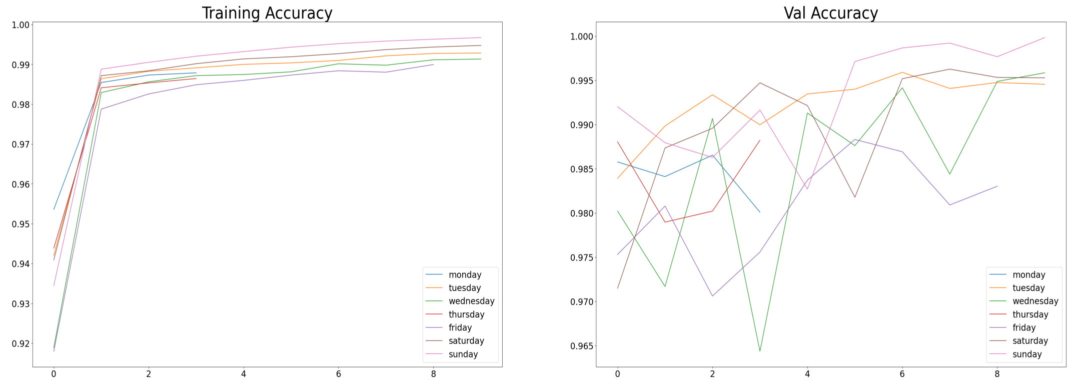


Figure 11: Training and val accuracies for each day model.

4.1.2 One model for all days in once training

The second approach in training is one model for all days in once. In this approach, model is trained whole days data of course with 256 batch size. It is trained with same hyperparameters of first approach. It tooks 12 minutes 38 seconds in total. Figure 12 shows the training and validaiton losses of this approach model. Also, figure 13 shows the training and validation accuracies for one model for all data approach.

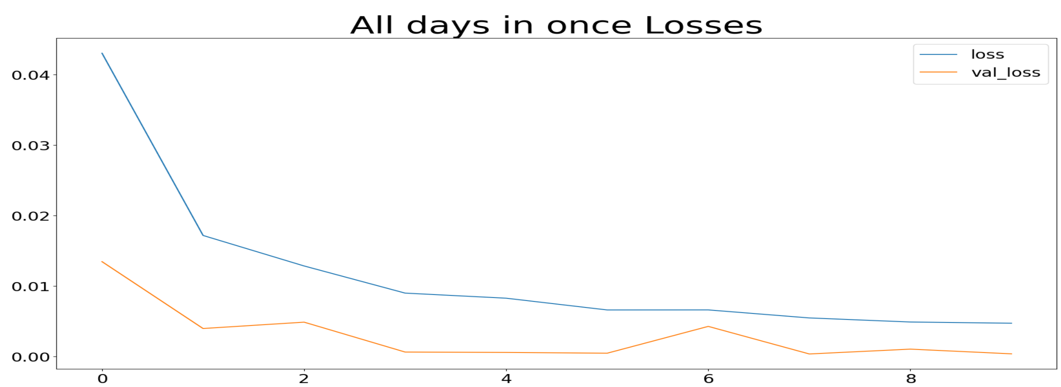


Figure 12: Training and val losses for one model for all data model.

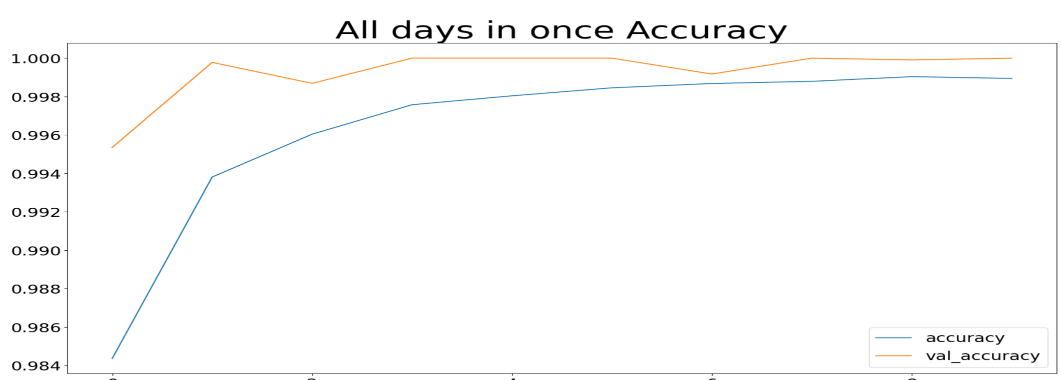


Figure 13: Training and val accuracies for one model for all data model.

5 Classification reports for two training approaches

Figure 14 shows the classification report for first approach which trained each model for each day. If we take the average of each day's weight avg, we obtain that approximately %99 f1 score which combines precision and recall. This is an encouraging result to use DeepLevel for predicting mobile traffic levels of urban area in real life applications.

MONDAY	precision	recall	f1-score	support
1	1.0000	0.9609	0.9801	53904
2	0.9490	0.9710	0.9599	40355
3	0.9599	0.9891	0.9743	28799
4	0.9680	0.9976	0.9826	20503
5	0.9783	0.9685	0.9734	13403
6	0.9990	0.9957	0.9973	59563
accuracy			0.9801	216527
macro avg	0.9757	0.9805	0.9779	216527
weighted avg	0.9805	0.9801	0.9801	216527

[a]

TUESDAY	precision	recall	f1-score	support
1	1.0000	0.9871	0.9935	54431
2	0.9821	0.9966	0.9893	39345
3	0.9951	0.9955	0.9953	29024
4	0.9929	0.9942	0.9935	20535
5	0.9917	0.9913	0.9915	13313
6	0.9986	1.0000	0.9993	60571
accuracy			0.9945	217219
macro avg	0.9934	0.9941	0.9937	217219
weighted avg	0.9945	0.9945	0.9945	217219

[b]

WEDNESDAY	precision	recall	f1-score	support
1	0.9992	0.9993	0.9992	54684
2	0.9984	0.9987	0.9986	39161
3	0.9928	0.9991	0.9960	29179
4	0.9996	0.9844	0.9919	20446
5	0.9915	0.9603	0.9756	13843
6	0.9911	1.0000	0.9955	60388
accuracy			0.9955	217701
macro avg	0.9954	0.9903	0.9928	217701
weighted avg	0.9955	0.9955	0.9955	217701

[c]

THURSDAY	precision	recall	f1-score	support
1	0.9935	0.9975	0.9955	54555
2	0.9930	0.9862	0.9896	40131
3	0.9884	0.9743	0.9813	29204
4	0.9679	0.9669	0.9674	20384
5	0.9536	0.9784	0.9658	13543
6	0.9958	0.9982	0.9970	59522
accuracy			0.9884	217339
macro avg	0.9820	0.9836	0.9828	217339
weighted avg	0.9885	0.9884	0.9884	217339

[d]

FRIDAY	precision	recall	f1-score	support
1	0.9857	0.9999	0.9928	53286
2	0.9999	0.9560	0.9774	39616
3	0.9670	0.9597	0.9634	29707
4	0.9446	0.9790	0.9615	20933
5	0.9563	0.9983	0.9769	14336
6	0.9997	0.9965	0.9981	61304
accuracy			0.9835	219182
macro avg	0.9755	0.9816	0.9783	219182
weighted avg	0.9838	0.9835	0.9835	219182

[e]

SATURDAY	precision	recall	f1-score	support
1	0.9878	1.0000	0.9939	59886
2	0.9959	0.9820	0.9889	41057
3	0.9953	0.9940	0.9947	28337
4	0.9993	0.9932	0.9963	19696
5	0.9993	0.9993	0.9993	13785
6	1.0000	0.9998	0.9999	53843
accuracy			0.9951	216604
macro avg	0.9963	0.9947	0.9955	216604
weighted avg	0.9951	0.9951	0.9951	216604

[f]

SUNDAY	precision	recall	f1-score	support
1	1.0000	0.9999	0.9999	59799
2	0.9999	0.9998	0.9998	40137
3	0.9997	0.9999	0.9998	27871
4	0.9998	0.9992	0.9995	19200
5	0.9986	1.0000	0.9993	13218
6	1.0000	0.9999	1.0000	50168
accuracy			0.9998	210393
macro avg	0.9997	0.9998	0.9997	210393
weighted avg	0.9998	0.9998	0.9998	210393

[g]

Figure 14: Classification reports for each model trained with each day.

Figure 15 shows the classification report for second approach which uses a one model for all days data in once. It has a very high f1 score as in first approach. Like in first approach, we can use this method of DeepLevel for mobile traffic levels of urban area in real life apps.

	precision	recall	f1-score	support
1	1.0000	1.0000	1.0000	391459
2	1.0000	1.0000	1.0000	279378
3	1.0000	1.0000	1.0000	201300
4	1.0000	1.0000	1.0000	141213
5	1.0000	1.0000	1.0000	96185
6	1.0000	1.0000	1.0000	405430
accuracy			1.0000	1514965
macro avg	1.0000	1.0000	1.0000	1514965
weighted avg	1.0000	1.0000	1.0000	1514965

Figure 15: Classification reports for one model trained with all day data in once.

6 Conclusion

In this article, we proposed DeepLevel model to predict mobility traffic levels in urban areas. We think that mobile traffic level prediction can be a key part of upcoming mobile technologies such as 5G network slicing. After research and training results, we suggest to you use one of two approaches which are described in article with respect to your dataset and needs. Both of two approaches are successful and robust in this big dataset as shown above results. Here, you can find the source code of DeepLevel. Of course, you absolutely free to change hyperparameters or all architecture of neural network.

References

[Ita15] Telecom Italia. Telecommunications - SMS, Call, Internet - MI, 2015.

A. Thantharate, R. Paropkari, V. Walunj and C. Beard, ”DeepSlice: A Deep Learning Approach towards an Efficient and Reliable Network Slicing in 5G Networks,” 2019 IEEE 10th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON), New York, NY, USA, 2019, pp. 0762-0767, doi: 10.1109/UEMCON47517.2019.8993066.

Barlacchi, G., De Nadai, M., Larcher, R. et al. A multi-source dataset of urban life in the city of Milan and the Province of Trentino. Sci Data 2, 150055 (2015). https://doi.org/10.1038/sdata.2015.55