JURNAL ILMIAH TEKNOLOGI INFORMASI DAN KOMUNIKASI (JTIK) VOL 13, No.2, September 2022, pp 102 – 131
p-ISSN : 2087-0868 (media cetak)
e-ISSN : 2598-9707 (media online)
http://ejurnal.provisi.ac.id/index.php/JTIKP

5G NETWORK TRAFFIC FORECASTING USING MACHINE LEARNING

Budi Raharjo[1], Mars Caroline Wibowo[2]
[1] University of Science and Computer Technology (STEKOM University), Semarang, 57166, Indonesia
e-mail: budiraharjo@stekom.ac.id
[2] University of Science and Computer Technology (STEKOM University), Semarang, 57166, Indonesia
e-mail: caroline@stekom.ac.id

ARTICLE INFO

Article history:

Received : 22–Maret-2022
Received in revised form : 25–April-2022
Accepted : 29–Juni-2022
Available online : 30–September-2022

ABSTRACT

The idea of network chunks being described as virtual subsets of the physical resources of 5G infrastructure is used in standards for 5G communications. The efficiency of ML predictors for traffic prediction in 5G networks has been established in recent research so that it becomes to assess the capability demands of each network slice and to see how it progresses as a large number of network slices are deployed over a 5G network over time to be very important. The main objective of this research is to establish the model that has the potential to help network management and resource allocation in 5G networks with machine learning performance analysis in predicting network traffic on high-dimensional spatial-temporal cellular data, in addition to investigating the effectiveness of various neural network models in traffic prediction from univariate and multivariate perspectives. The research method used is a quantitative research method using correlation analysis, statistical analysis, and distribution analysis on the temporal and spatiotemporal frameworks developed to predict traffic from a univariate and multivariate perspective. To predict 24-hour mobile traffic requires combining spatial and temporal dependencies. The univariate analysis will be carried out by applying a temporal framework that includes FCSN, 1DCNN, SSLSTM and ARLSTM to capture temporal dependencies. The results of various experiments in this study show that the proposed spatiotemporal model outperforms the temporal model and other techniques in the mobile traffic forecasting literature including internet, SMS, and calls. Therefore, it is expected that network optimization and more effective resource allocation can be carried out by predicting cellular traffic through the proposed 2D-ConvLSTM model. In future work, prediction performance can be improved and 2D-ConvLSTM model deployment in 5G networks can be optimized automatically

Keywords: *5G network, machine learning network, traffic forecasting, spatiotemporal models, temporal analysis.*

Received: 4–Maret-2022; Revised 5–April-2022; Accepted 1–September-2022

p-ISSN : 2087-0868 e-ISSN : 2598-9707

INTRODUCTION

As cellular technology advances toward the current fifth-generation (5G) technology, telecommunication traffic forecasting becomes very important, especially in terms of providing intelligent management features. “Assessing the resources on each network slice and analyzing changes in a large number of network chunks over time deployed over a 5G network is critical” (Ravindran, et al 2016). Network management and optimization strategies are more effective when mobile operators know in advance the demand for mobile data traffic at specific times and locations. Because the accuracy of prediction models is of great importance, this study focuses on near-term mobile traffic predictions using temporal and Spatio-temporal approaches. A real cellular traffic dataset was used in this study to evaluate the performance of various neural network models in predicting cellular traffic. The proposed Spatio-temporal network can be practically used for resource allocation and network management.

5G networks use a variety of technologies to meet these requirements, network cutting being one of the most important. Shared network resources’ physics can be actively and effectively scattered into intelligent network slices depending on changing user needs by leveraging cloud computing and virtualization technologies. It is very important to identify the resource demands and these requirements change supplementary. If a network slice requires more resources than originally allocated, it is considered under-available. This results in poor network performance and QoS for users. Conversely, if a network slice uses fewer resources, it will be over-provisioned. In this case, the resource is not required but the resource will remain active, this can incur expenses on the infrastructure provider. Dynamically adjusting the resources allocated to network slices is critical as both scenarios impose costs on the infrastructure provider and lead to a reduction in the quality of service. Therefore, dynamically allocating resources in recognition of the traffic profile of each slice becomes very important. In addition, with the development of 5G technology, communication networks are becoming smarter and self-organizing. Afolabi (2017) explained “ML and time series analysis, which have been used in various applications, are considered powerful tools for modeling and forecasting network traffic. Incorporating adaptable machine intelligence into future cellular networks is attracting much attention in the scientific community”. According to the research by Li, et al 2014 and Hu, et al 2015 described “This trend is represented in the development of network systems that utilize machine learning techniques to address challenges ranging from ‘*radio access technology or* ‘RAT’ selection to malware diagnosis”.

p-ISSN : 2087-0868 e-ISSN : 2598-9707

ML can be used to systematically extract appropriate information from movement data and automatically detect relationships too complex for human experts. In this study, the predictive performance of several neural network models is evaluated in mobile traffic forecasting. In particular, temporal and spatiotemporal frameworks were developed for univariate and multivariate analysis to predict mobile traffic 24 hours into the future.

The data generated by mobile devices varies widely because it is often aggregated in multiple formats from multiple sources. Classical machine learning methods are rendered unfeasible for solving challenges in this space such as data abstraction and meaning. According to the research by Chen et al, 2015, "as the data scale increases, performance does not increase". Research by Kazmier, 2004 explained, "in control problems, they cannot manage highly dimensional spaces". Thus, the integration of deep learning into 5G cellular and wireless networks is fully justified by automated and intelligent data representation and feature selection. This means that data can be effectively filtered and higher-level abstractions detected while reducing the need for preprocessing. The research by Wang et al, 2017 "proposed an approach that combines auto-encoders with '*long short-term memory'/*'LSTM' networks to take advantage of the spatial dependence of different cells". However, by Geoffrey (2006) explained about "the representations learned by auto-encoders are missing depictions of the original data and they may not adequately capture the spatial dependence of nearby cells".

Data Sets

The multi-source dataset generated by Indonesian telecommunications end of 2017 is used in this research. This is one of the most comprehensive assortments of operators. This collection was originally developed to address big data challenges with approaches ranging from mobile networks to communal utilization. This data set consists of data archives regarding information technology, climate, disclosure, social networks, and utilities from 2013 to 2017. In this study, telecommunications records are used to predict traffic. The main target of this research is to establish a model that has the potential to help network management and resource allocation in 5G networks with machine learning performance analysis in predicting network traffic on high-dimensional spatial-temporal cellular data, in addition to investigating the effectiveness of various neural network models in traffic prediction from univariate and multivariate perspectives. The univariate analysis will be carried out by applying a temporal framework that includes a "*Fully connected sequential Network" "*FCSN", "*one-dimensional convolutional neural network" "1DCNN*", "*single-shot learning LSTM" "*SSLSTM", and "*autoregressive LSTM" "*ARLSTM" to capture

p-ISSN : 2087-0868 e-ISSN : 2598-9707

temporal dependencies. In the second part, a multivariate spatiotemporal analysis will be performed using a 2-dimensional convolution LSTM (2D-ConvLSTM) to forecast traffic on Indonesian telecommunication data. The goal of the multivariate spatiotemporal analysis is to incorporate dependencies between different variables, and spatial and temporal information into predictive modeling automatically.

LITERATURE REVIEW

5G is a more powerful unified air interface designed with expanded capacity to enable next-generation user experiences, enable new delivery models, and deliver new services. 5G wireless technology is expected to deliver higher data rates, ultra-low latency, higher reliability, massive network capacity, improved availability, and a more consistent user experience for more users. Higher performance and increased efficiency will enable new user experiences and connect new industries. The industry is considering seeing how the network can be used to overcome the intense capacity and conduct demands in the future as the demand for enhanced mobile broadband experiences continues to grow. Enabling multiple platforms to function together as a unified entity, mostly controlled by software and adaptable to any consumption pattern, is the real challenge. In this context, 5G is expected to meet industrial and social demands. It focuses on increasing capacity by combining existing methodologies with advances in radio technology or if necessary, transformations in system design concepts. In addition, 5G has provided fast and wide internet coverage since 2020. A standardized and more uniform solution will enable much greater volume and therefore higher integration density. In addition, the proposed solution will lower energy use, and reduce costs. The key enabler of 5G networks is the use of *artificial intelligence* (AI).

METHODOLOGY

The temporal and spatiotemporal frameworks were developed to predict traffic from univariate and multivariate perspectives. Within the temporal framework, “FCSN”, “1DCNN”, “SSLSTM”, and “ARLSTM” are used to predict mobile traffic in SMS, call, and internet time series individually. To incorporate the dependence of the spatial and temporal data, a multivariate analysis was also performed. For the spatiotemporal framework, a 2-dimensional Convolution LSTM model is proposed to predict 24-hour mobile traffic using multi-channel data including SMS, call, internet, and count. Next, to evaluate the supposed model’s performance for both frameworks, the basic model is given here.

p-ISSN : 2087-0868 e-ISSN : 2598-9707

The appropriate files for investigation are collected by organizing the original data set according to "DateTime" and "square". In each “Grid ID”, one sub-set is created in a total of 10,000 subsets. Besides a large number of traffic cells is zero in ten minutes timeframe of the data set, which creates the data very sparsely. Furthermore, resource planning at the 10-minute level is a difficult task that can lead to network instability or intensive aerial. Therefore, the data is resampled every hour by totaling the traffic. The 3 parts that are added to the data file are "count", "sms" and "calls".

	squareid	internet	count	sms	calls	weekend	holiday	part_of_day
squareid	1.000000	0.134045	0.111384	0.114984	0.110346	-0.000003	0.000103	-0.000087
internet	0.134045	1.000000	0.442599	0.872689	0.835118	-0.034976	-0.043935	0.100908
count	0.111384	0.442599	1.000000	0.490132	0.512058	-0.114089	-0.039152	0.253803
sms	0.114984	0.872689	0.490132	1.000000	0.917543	-0.070013	-0.020092	0.134213
calls	0.110346	0.835118	0.512058	0.917543	1.000000	-0.088326	-0.052294	0.136981
weekend	-0.000003	-0.034976	-0.114089	-0.070013	-0.088326	1.000000	-0.082098	0.000410
holiday	0.000103	-0.043935	-0.039152	-0.020092	-0.052294	-0.082098	1.000000	0.010747
part_of_day	-0.000087	0.100908	0.253803	0.134213	0.136981	0.000410	0.010747	1.000000

Figure 1 – Interaction (correlation) between variables

The total of days of the week from the "datetime" column is extracted. Weekends are separated from weekdays which means Saturdays and Sundays are marked True and weekdays False in the “weekend” column. Next, it takes the holidays in November and December, and the first day of January and generates a “vacation” feature. Additionally, a new feature called "sections of the day" where the hours between 06:00 and 18:00 is considered the time of day indicated by 1 and the time of night by 0.

Correlation analysis

Next, the correlations between the different variables are calculated, as shown in Figure 1, with correlation coefficients in the range of -1 to 1. The coefficient of nearly 1 indicates a significant and specific relation between the two variables, implying that as one grows, the other will increases, and if one decreases, the other also decreases. A coefficient of around -1 indicates a significant negative relationship between the two variables, indicating that observations with high values in one variable tend to have lower values in the other variable or vice versa. Additionally, if the coefficients are close to zero there is no linear relationship between the two variables. Figure 1 shows the positive correlation between "sms", "internet", & "calls". Also, there is a slight correlation between "day parts", "count", "internet", "sms" and "calls".

p-ISSN : 2087-0868 e-ISSN : 2598-9707

Statistic analysis

Also, the skewness and kurtosis of all features are calculated as shown in Table 2. Because the skewness of all features is positive, the features have a long right tail. In addition, because the internet, count, call, and sms kurtosis is greater than 3, this indicates that the dataset has heavier tails than the normal distribution illustrated in Figure 2.

Table 1 - Slopes and kurtosis of features

Feature	Skewness	Kurtosis
Internet	7.37	102.40
Count	1.78	6.06
Call	7.96	111.51
SMS	8.97	117.78

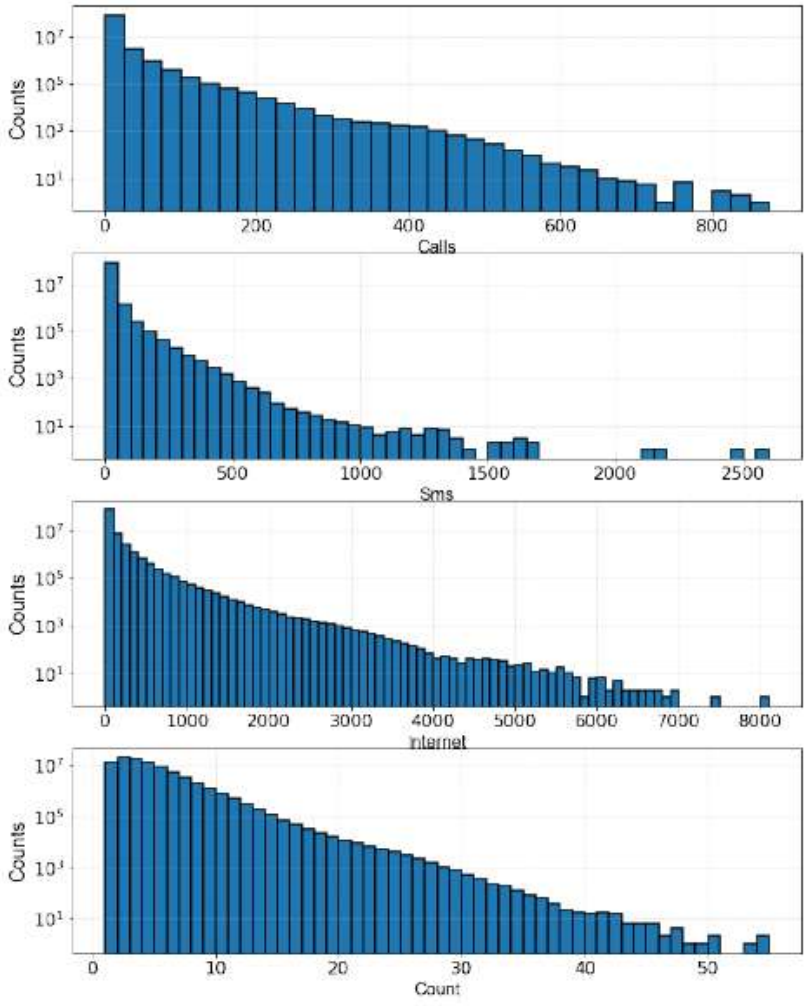


Figure 2 - Skewness and kurtosis of features

In addition, for a visual assessment of any abnormality and any correlation between the data recorded on holidays and weekends, the time series of the recorded data with their corresponding means and standard deviations were plotted.

p-ISSN : 2087-0868 e-ISSN : 2598-9707

Distribution analysis

Furthermore, the sms, count, call, and internet probability density functions (PDF) for “grid 5161” and “grid 7524” are illustrated in Figure 5. For the "5161 grid" the distribution is bimodal which has two peaks. Whereas on the "7524 grid", because there is a sudden increase on certain days bimodality is not observed.

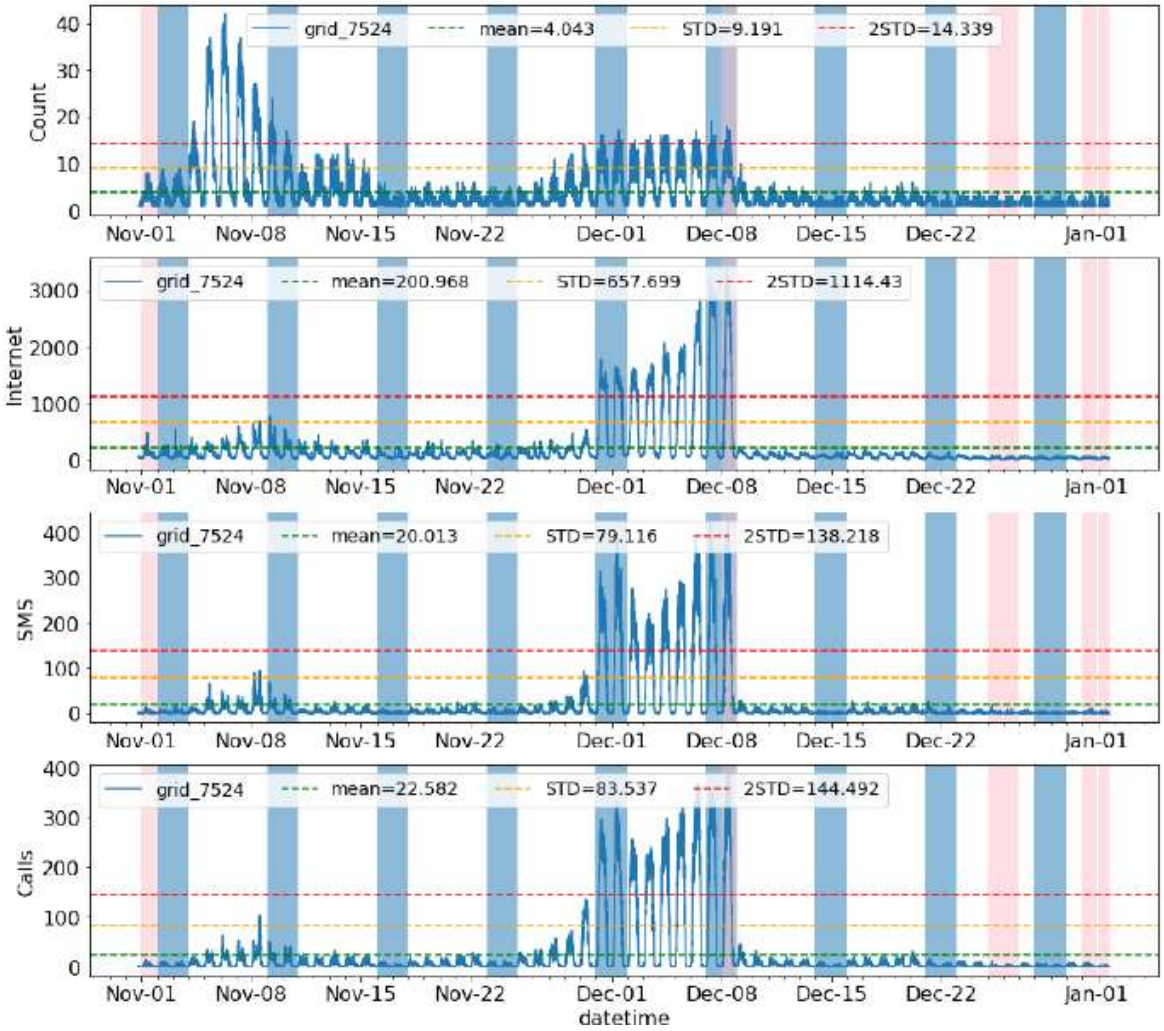


Figure 3 - Time series visualization of the '7524 grid'

p-ISSN : 2087-0868 e-ISSN : 2598-9707

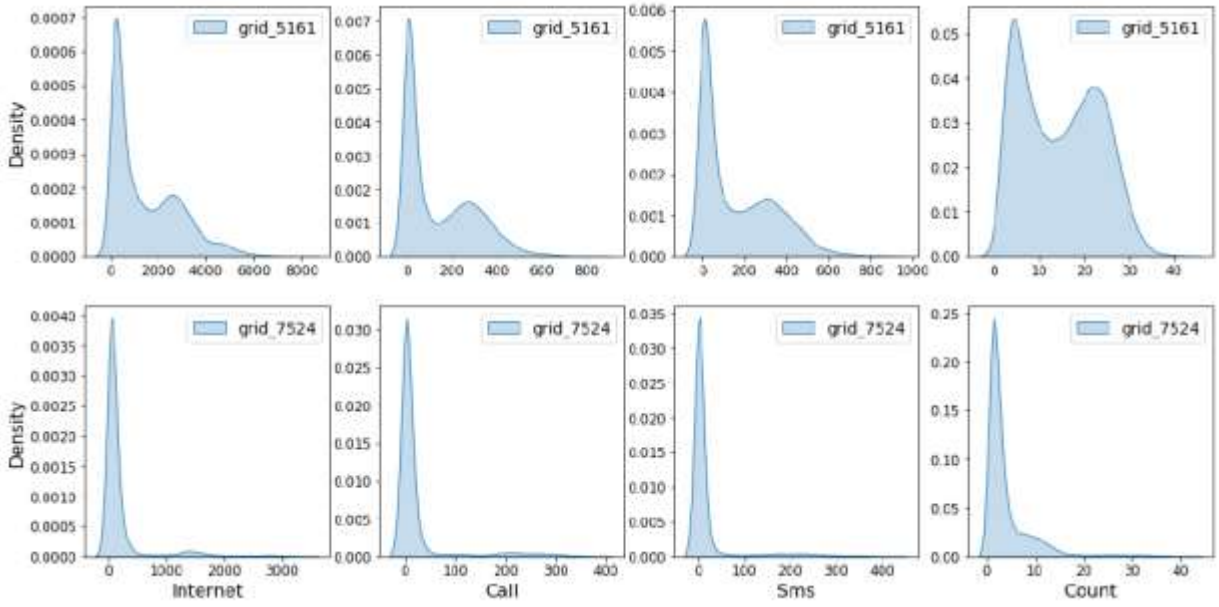


Figure 4 - PDF of 'grid 5161" and "grid 7524"

For further analysis, a data frame for internet, sms, and call averages for all network IDs, a feature histogram, based on their averages, is generated and displayed with a right-slant.

Temporary data set

Since datetime as a string is useless, the "datetime" value is changed to insufficient. Nonetheless, the data has daily and weekly periodization. To work around this problem, the time-frequency representation is as follows:

$$f^s(\xi) = \sin\frac{2\pi t}{P}$$

$$f^s(\xi) = \cos\frac{2\pi t}{P}$$

where

t = *time in second*

P = *cycle length*

In this temporal dataset, "internet", "count", "sms", "calls", "Day sin", "Day cos", "week sin" and "week cos" are used as inputs to predict mobile traffic from each grid for the next 24 hours features internet, count, sms, call, and frequency based on the previous 24 hours.

p-ISSN : 2087-0868 e-ISSN : 2598-9707

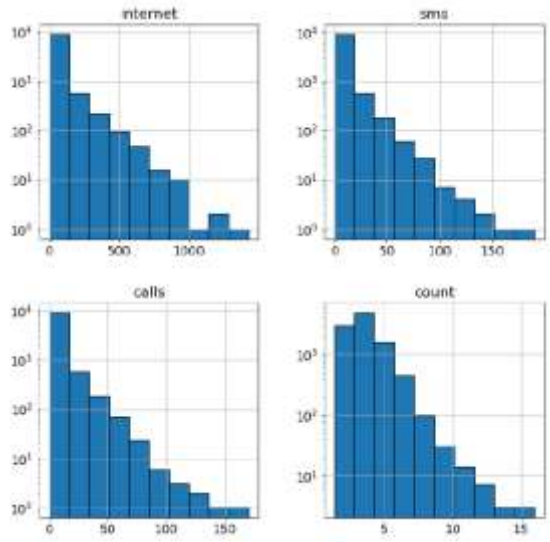


Figure 5 The average traffic of each network

Spatiotemporal data set

In a spatiotemporal dataset, the observed data for a given time is like a frame with 100 × 100 pixels. The spatiotemporal dataset looks like (1487,100, 100, 4) where 1487 denotes the time step (hours), 100 × 100 denotes longitude and latitude and 4 denotes the channel including internet, count, sms, and call as model inputs. In addition, Min-Max scaling is applied to the data set to re-scale the range of features between the ranges in [0, 1]. In Figure 6, an x grid and a y grid containing 100 × 100 grids (10,000 grids in total), and time steps from 1 to 1487 hours, are illustrated. In addition, on the right side is shown the frame of the last time step where each cube contains internet, count, sms, and call records. Higher internet consumption is shown in a lighter color while low internet usage is depicted in a darker color.

Interim predictive models

Several neural network-based models, as well as base models, are used.

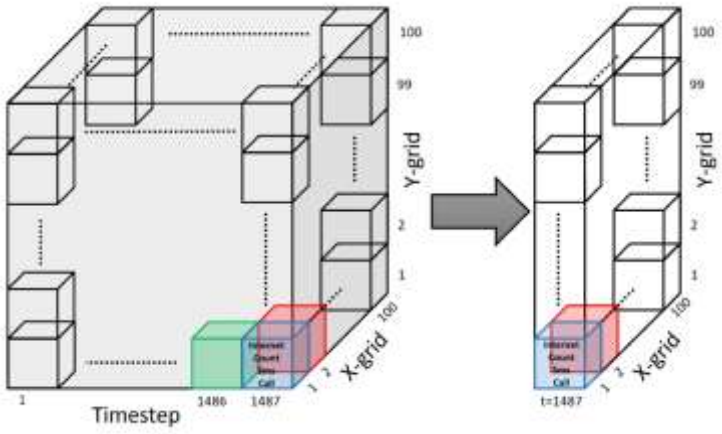


Figure 6 - Spatiotemporal data schema

p-ISSN : 2087-0868 e-ISSN : 2598-9707

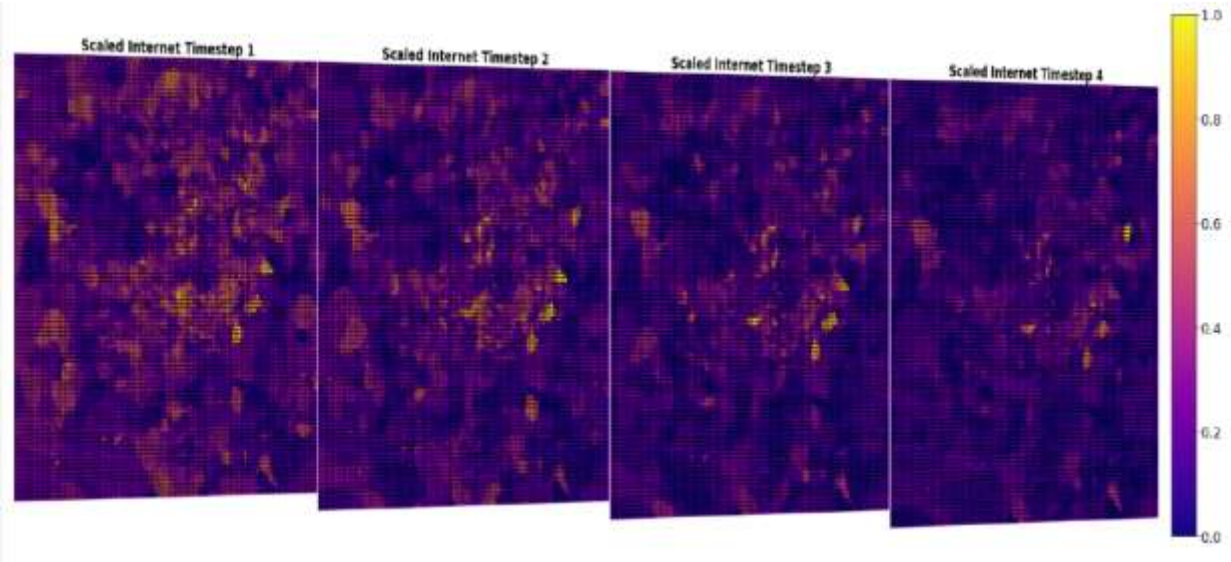


Figure 7 - Internet usage frames from time steps 1 to 4.

The temporal basis predictive model

The work's models of machine learning are very useful for establishing a baseline for comparison. As a temporal baseline, it is assumed that the 24-hour prediction will have the same pattern.

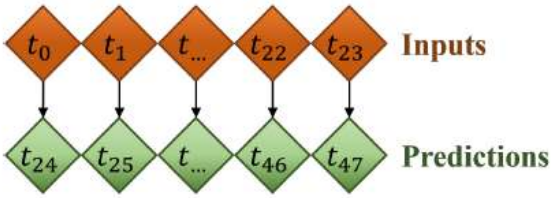


Figure 8 -Baseline Structure Block Diagram.

FCSN

According to the research by Volodymyr, et al (2015) "A fully connected neural network consists of a sequence of fully connected layers, where each layer's neurons are connected to neurons in another layer, a crucial advantage of quite united networks is that they are 'structure agnostic,' sense no specific hypo-research on input is required. This study proposes an FCSN that stacks two dense layers in Figure 9. "R*ectified linear unit activation"* or "RELU" is accustomed study complex patterns in data. Also, the graph in Figure 9 illustrates the connections in a neural network. Each node in this diagram is labeled with the shape of its input and output matrix.

p-ISSN : 2087-0868 e-ISSN : 2598-9707

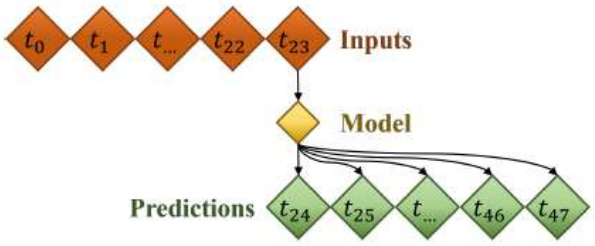


Figure 9 - FCSN Structure Block Diagram.

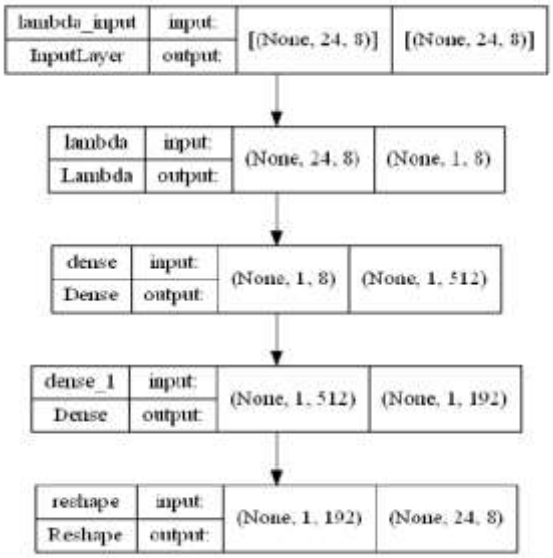


Figure 10 - Graphical visualization of the FCSN model.

1DCNN

A one-spatial involutional network produces an output tensor by connecting an input with a one-spatial involutional kernel. A convolutional model makes predictions based on a fixed-width history. It can perform better than dense models because it allows you to observe changes over time. In this study, his 1DCNN with kernel size 6 and RELU activation function was developed to predict communication traffic for the next 24 hours. This study also uses Graphviz to describe connections between neural networks. Each node in this diagram is labeled in the form of the input and output matrix.

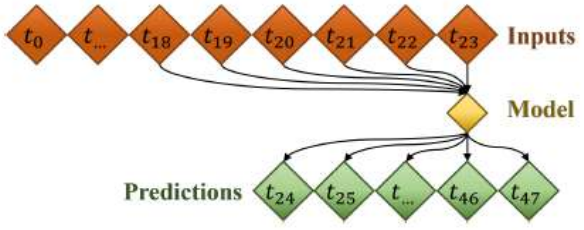


Figure 11 - 1DCNN Structure Block Diagram.

p-ISSN : 2087-0868 e-ISSN : 2598-9707

SSLSTM

LSTM networks are a special type of RNN developed to prevent the missing gradient problem. By utilizing LSTM cells long-term dependence can be learned from the data (Hochreiter , et al 1997). The structure of the LSTM unit makes it possible to study long-term dependencies. Unlike ordinary neurons, LSTMs contain gates that control the learning process. Through the use of structures known as gates, each LSTM cell can retain or forget information about previous network states. Additionally, LSTM has proven its prowess in speech translation. The model gathers for 24 hours of internal status before providing a single estimate for the next 24 hours.

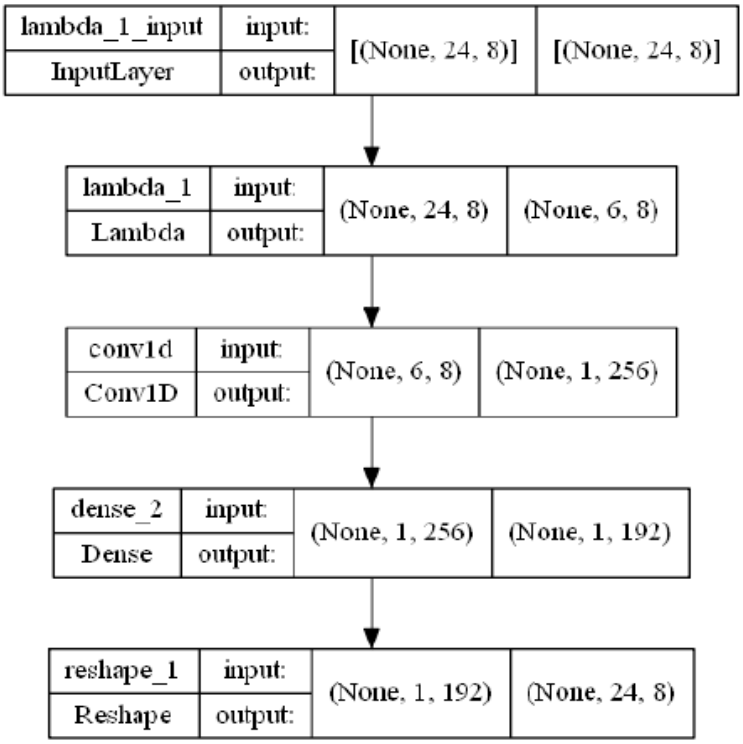


Figure 12 - Graphical visualization of the 1DCNN model.

The graph in Figure 13 shows the connections in the LSTM network. Each node in this diagram is labeled with the shape of its input and output matrix.

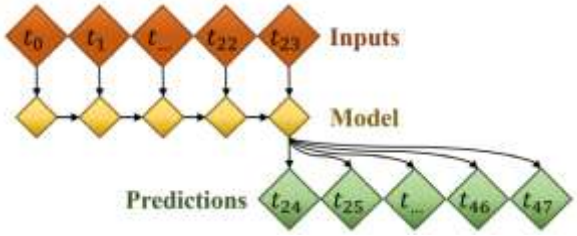


Figure 13 - SSLSTM Structure block diagram.

p-ISSN : 2087-0868 e-ISSN : 2598-9707

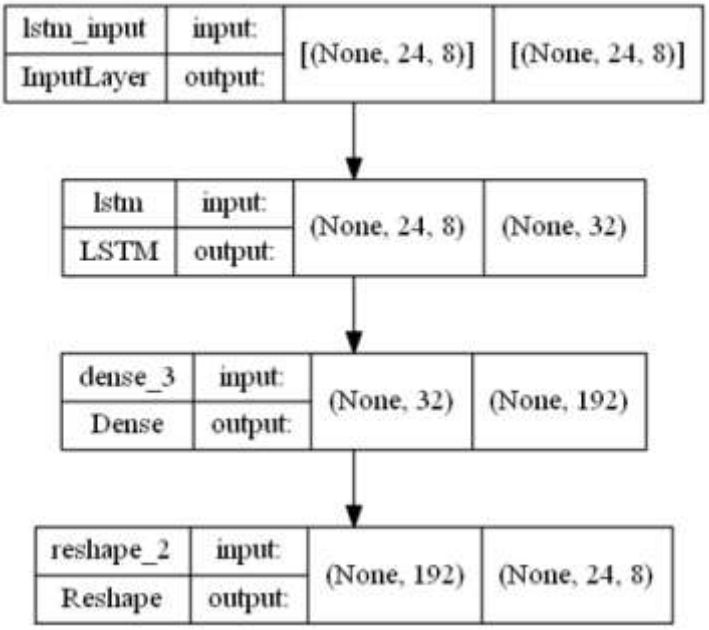


Figure 14 - Graphical visualization of the SSLSTM model.

ARLSTM

All the above models give the integrated output sequence in one step. It may be useful for the model to decompose these predictions into discrete times. As in a classical's RNN, each model's output can be fed back to itself at each stage to generate a sequence and make predictions based on the previous model. The position of this type of model is that it can be tuned to produce outputs of different lengths. In this work, a sequential model is proposed by packing the LSTM cell layer into the lower level of the RNN layer to simplify the "warm-up" method of predicting communication service for the next 24 hours. The warm-up method displays single-time step predictions and the internal state of the LSTM. With the state of the RNN and initial predictions, it is possible to continue with the iteration of the model and enter the predictions at each time step as input to predict the next time step. In addition, a summary of the ARLSTM model is illustrated in Figure 16. The model summary contains information, output format, total parameters and total in each layer.

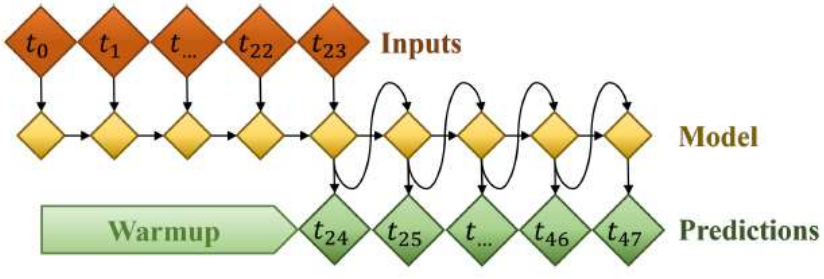


Figure 15 - ARLSTM Structure Block Diagram.

p-ISSN : 2087-0868 e-ISSN : 2598-9707

Spatiotemporal models

The model of 2D ConvLSTM is suggested for the multivariate Spatio-temporal investigation to predict mobile traffic in the next 24 hours, moreover, the spatiotemporal baseline is used as a benchmark to evaluate the effectiveness of the 2D-ConvLSTM model.

The basic spatiotemporal model

Before developing a spatiotemporal model, it is helpful to establish a performance baseline for comparison.

```
Model: "AR-LSTM"
_________________________________________________________________
 Layer (type)                 Output Shape              Param #
=================================================================
 lstm_cell_1 (LSTMCell)       multiple                  5248

 rnn (RNN)                    multiple                  5248

 dense_6 (Dense)              multiple                  264

=================================================================
Total params: 5,512
Trainable params: 5,512
Non-trainable params: 0
_________________________________________________________________
None
```

Figure 16 - Summary of the ARLSTM model.

2D-ConvLSTM

To extract the dependencies of spatial and temporal data simultaneously and incorporate them in traffic predictions, a 2D-ConvLSTM network is proposed to analyze multichannel spatiotemporal data. The proposed 2D-ConvLSTM framework consists of 4 Convolutional 2D layers of LSTM and a layer of 3D Convolutional. The proposed spatiotemporal model framework is illustrated in Figure 17. For the input, data in the form of (24,100,100.4) is considered capable of performing multivariate analysis and includes correlations between variables, space, and time. To be exact, the input data consists of 24-hour records across all networks for 4 channels including internet, count, sms, and call records. In the proposed model, the input data with the shape mentioned above is fed into the 2D Convolution LSTM layer which extracts the features in 10 channels. In the next layer, the 10 channels obtained are scaled down through progressive channel sorting to 8, 5, and 3 channels respectively, for optimization and reduction of the number of parameters. The features extracted in each layer are normalized for mean centering and variance scaling. This batch normalization reduces the risk of shifting the internal co-variate in the next layer, the 3 channels obtained are then fed into the 3D convolution layer in the last layer to predict the

p-ISSN : 2087-0868 e-ISSN : 2598-9707

next mobile traffic frame for each data type, d. D ∈ {internet, sms, call} denotes a particular type of mobile traffic that is estimated individually by the proposed spatiotemporal 2D-ConvLSTM model.

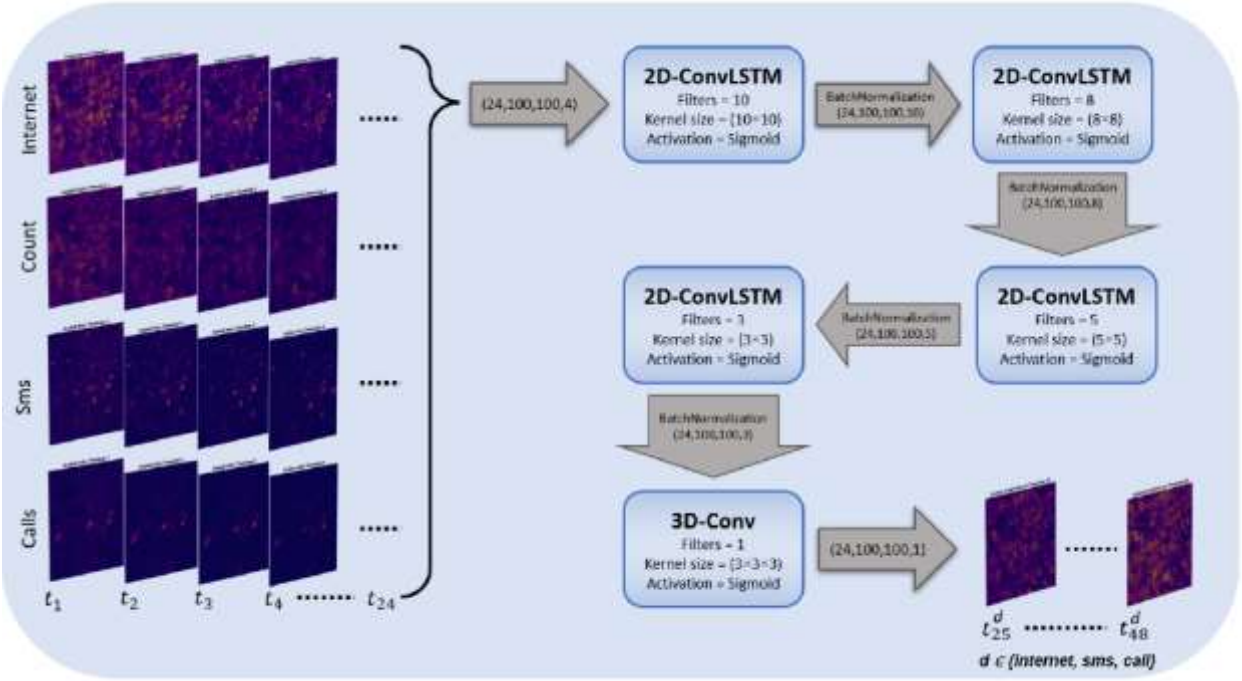


Figure 17 - The 2D-ConvLSTM model framework.

RESULTS AND EVALUATION

Evaluation metrics

To assess the work of the proposed prediction model correlated to another works, in this research the temporal and spatiotemporal models were evaluated by utilizing MAE and RMSE.

Absolute mean error (MAE)

The average size of the error in a series of predictions is measured by MAE. Absolute differences between model predictions and actual observations are averaged over the test sample with MAE.

$$MAE = \left(\frac{1}{n}\right)\sum_{i=1}^{n}|y_i - y_i|$$

where

n = *total number of data points*

yi = *actual output value*

yi = *predicted value for the ith data point.*

p-ISSN : 2087-0868 e-ISSN : 2598-9707

```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                 Output Shape              Param #
=================================================================
 conv_lstm2d_4 (ConvLSTM2D)  (None, 24, 100, 100, 10)  56040

 batch_normalization_4 (Batc  (None, 24, 100, 100, 10)  40
 hNormalization)

 conv_lstm2d_5 (ConvLSTM2D)  (None, 24, 100, 100, 8)   36896

 batch_normalization_5 (Batc  (None, 24, 100, 100, 8)  32
 hNormalization)

 conv_lstm2d_6 (ConvLSTM2D)  (None, 24, 100, 100, 5)   6520

 batch_normalization_6 (Batc  (None, 24, 100, 100, 5)  20
 hNormalization)

 conv_lstm2d_7 (ConvLSTM2D)  (None, 24, 100, 100, 3)   876

 batch_normalization_7 (Batc  (None, 24, 100, 100, 3)  12
 hNormalization)

 conv3d_1 (Conv3D)           (None, 24, 100, 100, 1)   82

=================================================================
Total params: 100,518
Trainable params: 100,466
Non-trainable params: 52
_________________________________________________________________
```

Figure 18 - Summary of the 2D-ConvLSTM model.

RMSE

RMSE or *"Root mean squared error"* is the equal root of the moderate of the marked characteristic between the predicted and observed values.

$$RMSE = \sqrt{\left(\frac{1}{n}\right)\sum_{i=1}^{n}(y_i - y_i)^2}$$

where

n = *total number of data points*

yi = *actual output value*

yi = *predicted value for the ith data point.*

RMSE is more sensitive to data with larger differences between actual and predicted values. RMSE is more sensitive to outliers compared to MAE. In fact, data with higher errors will skew the RMSE

p-ISSN : 2087-0868 e-ISSN : 2598-9707

Temporal mobile traffic prediction

The data set is allocated 70% in every grid, and every grid consists of 1487file. From 1040 records divided into 298 records for the validation set, 20% for the instructions set and 10% of the data set 149 records for corresponding, are devoted to the test set. In addition, the hyperparameters are set to control the learning process toward optimal predictions.

- Temporal baseline: The lowest validation loss was achieved by utilizing MSE as a loss function, Adam optimizer, and 0.001 learning rate. In addition, the temporal baseline is prepared based on the use of time step 1 to predict the first observation point, time step 2 to predict the 2nd observation point, and so on, up to the use of the 24th time step to predict traffic at the 24th observation point.

Table 2 - Temporal baseline hyperparameters

Criteria / Model	Loss		Optimizer			Learning rate		
	MAE	MSE	Adam	Adamax	SGD	0.1	0.01	0.001
Temporal baseline	x	✓	✓	x	x	x	x	✓
Model structure	• Considering time step 23 to predict the next 24 hours. • **Using the time step 1 to predict the first observed point, time step 2 for predicting the 2nd observed point and so on, until using time step 24 to predict the traffic of 24th observed point.**							

- FCSN: The FCSN model hyperparameters are illustrated in Table 3. The lowest loss validation was obtained by using MSE as a loss function, Adam optimizer, and a learning rate of 0.001. The best prediction results were obtained by utilizing RELU activation and 512 neurons in tbh 80he hidden layer.

Table 3 - FCSN Hyperparameters.

Criteria / Model	Loss		Optimizer			Learning rate		
	MAE	MSE	Adam	Adamax	SGD	0.1	0.01	0.001
FCSN	x	✓	✓	x	x	x	x	✓
Model structure	• Considering **RELU**, leaky RELU, and Tanh activation functions. • Considering 256, **512**, and 1024 units for hidden layer.							

- 1DCNN: Table 4 shows the hyperparameters used in the 1DCNN model. Using MSE as the loss function, Adam's optimizer, and a learning rate of 0.001, the lowest validation loss is observed. 1DCNN achieves the best results by leveraging a kernel size of 6, RELU activation function, and 256 filters in the Conv1D layer.

p-ISSN : 2087-0868 e-ISSN : 2598-9707

Table 4 - 1DCNN Hyperparameters.

Criteria / Model	Loss		Optimizer			Learning rate		
	MAE	MSE	Adam	Adamax	SGD	0.1	0.01	0.001
1D-CNN	✗	✓	✓	✗	✗	✗	✗	✓
Model structure	• Considering kernel size of 4, 5, **6**. • Considering **RELU**, Tanh, and sigmoid activation functions. • Considering 128, **256**, and 512 filters in Conv1D layer.							

- SSLSTM: Tuning the hyperparameters of the SSLSTM model is illustrated in Table 5. Validation loss has the lowest number when using MSE as a loss function, Adam optimizer, and 0.001 learning rate. In addition, the best results were obtained using 32 LSTM units and none dropped out.

Table 5 - SSLSTM Hyperparameters.

Criteria / Model	Loss		Optimizer			Learning rate		
	MAE	MSE	Adam	Adamax	SGD	0.1	0.01	0.001
SS-LSTM	✗	✓	✓	✗	✗	✗	✗	✓
Model structure	• Considering **32**, 64, 128, 256 LSTM units. • Considering **0**, 0.2 and 1 for the dropout.							

- ARLSTM: Table 6 illustrates that the lowest validation loss in the ARLSTM model is obtained by using MSE as a loss function, Adam optimizer, learning rate of 0.001, and 32 LSTM units.

Table 6 - ARLSTM Hyperparameters.

Criteria / Model	Loss		Optimizer			Learning rate		
	MAE	MSE	Adam	Adamax	SGD	0.1	0.01	0.001
SS-LSTM	✗	✓	✓	✗	✗	✗	✗	✓
Model structure	• Considering **32**, 64, 128, 256 LSTM units.							

Predictions with the baseline model in the "5161 grid" are illustrated in Figure 19, Figure 20, Figure 21, Figure 22, and Figure 23. In addition, the global work of the prediction plan is evaluated on the test set. FCSN and 1DCNN have proportionate work, but 1DCNN is a smaller system with fewer parameters.

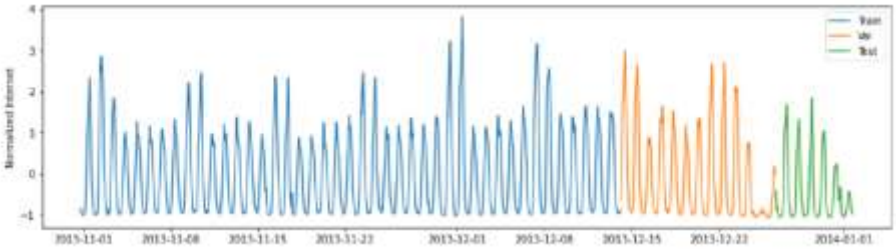


Figure 18 - Illustration of carriage sharing, validation, and test set for "grid 5161" normalized internet traffic

p-ISSN : 2087-0868 e-ISSN : 2598-9707

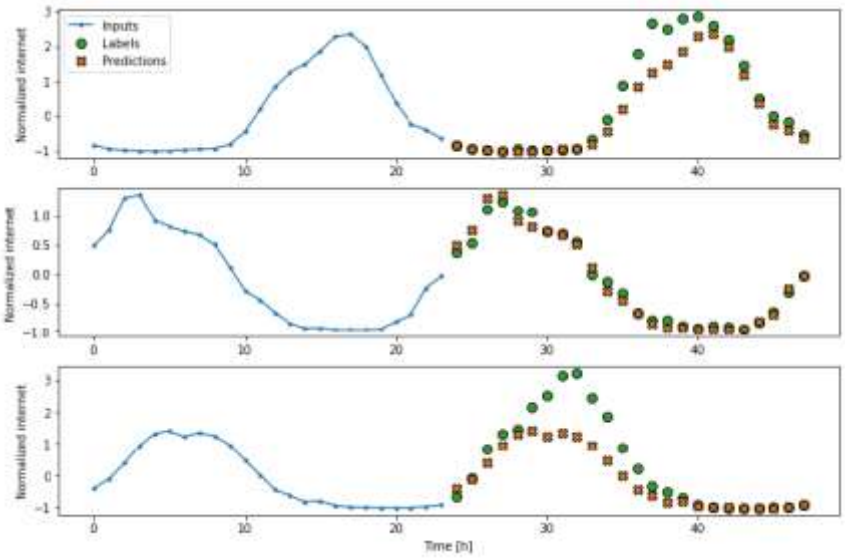


Figure 19 - Prediction of the base model on "grid 5161" normalized internet traffic.

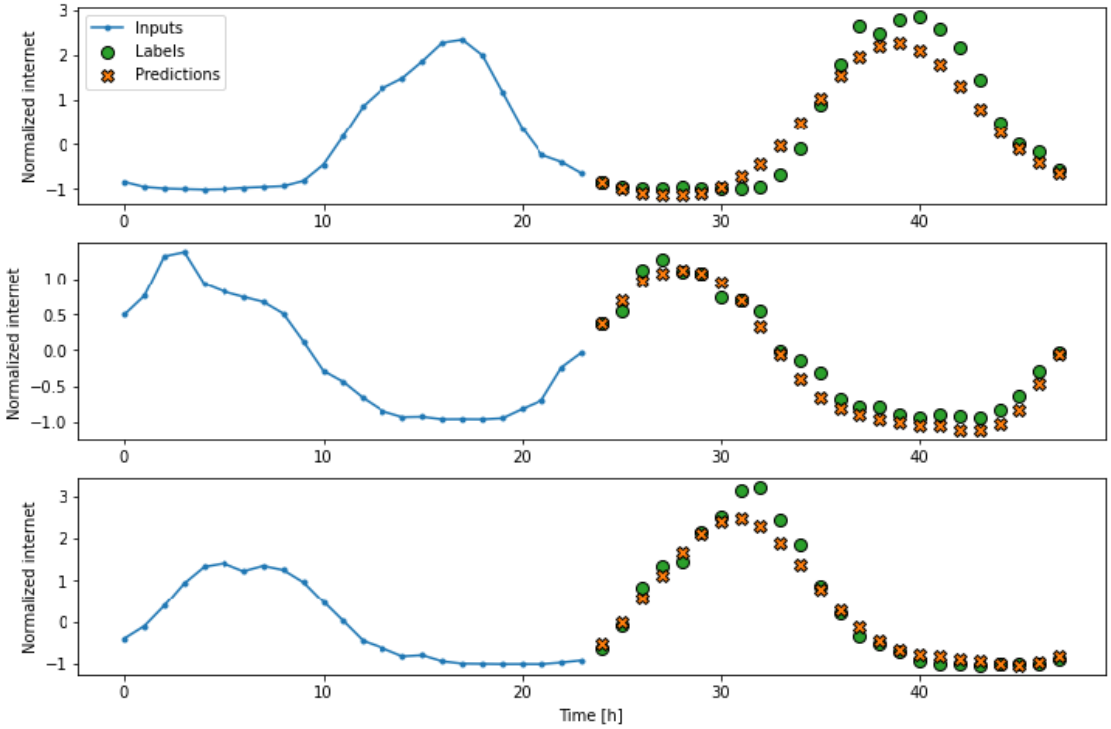


Figure 20 - FCSN model prediction on "5161 grid" normalized internet traffic.

p-ISSN : 2087-0868 e-ISSN : 2598-9707

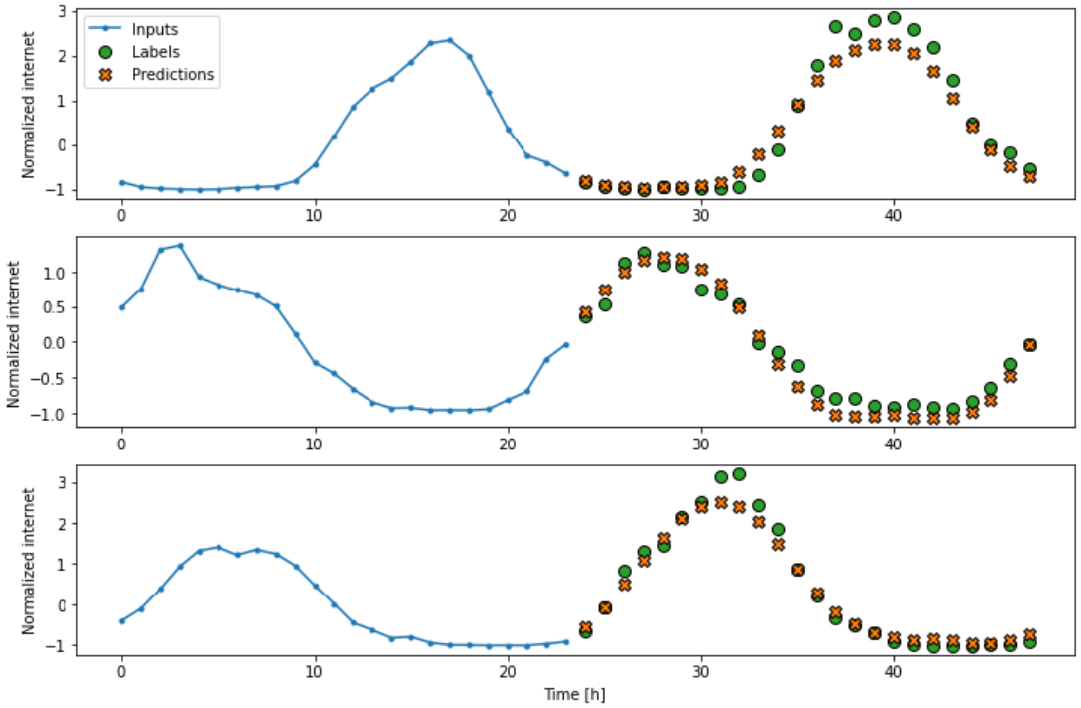


Figure 21 - 1DCNN model prediction on "grid 5161" normalized internet traffic.

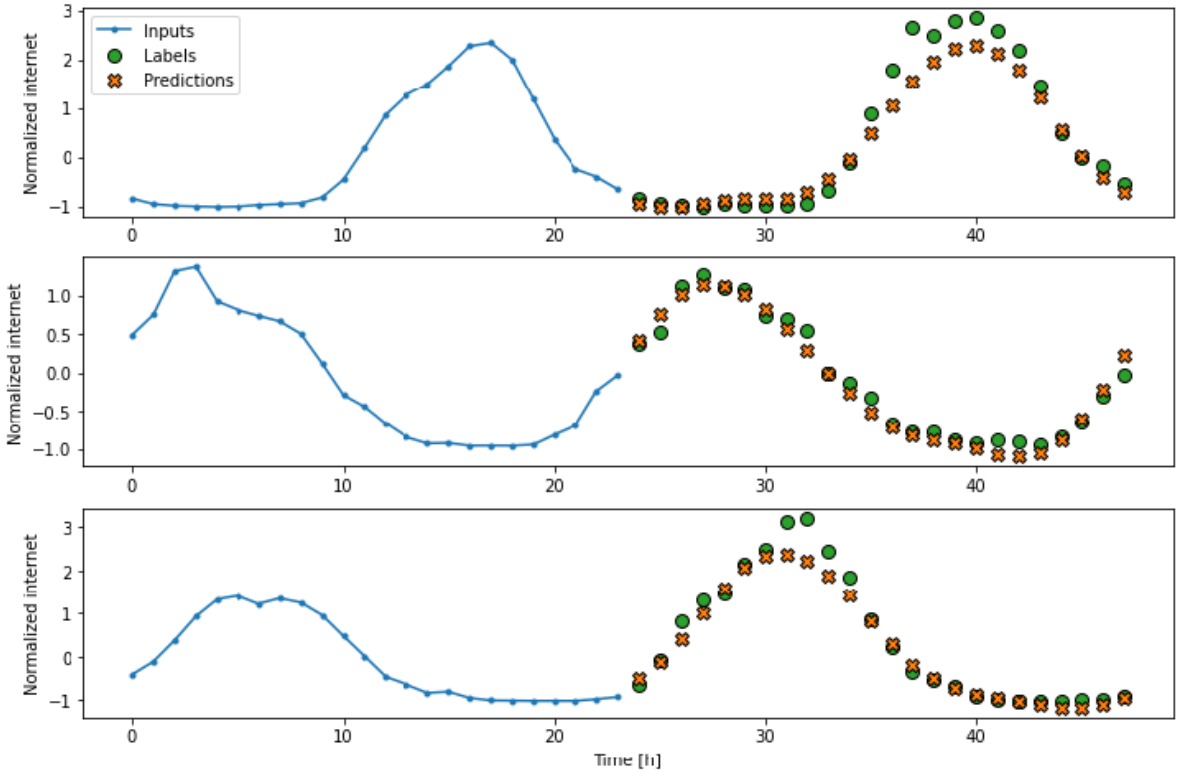


Figure 22 - Prediction of the SSLSTM model on "grid 5161" normalized internet traffic.

p-ISSN : 2087-0868 e-ISSN : 2598-9707

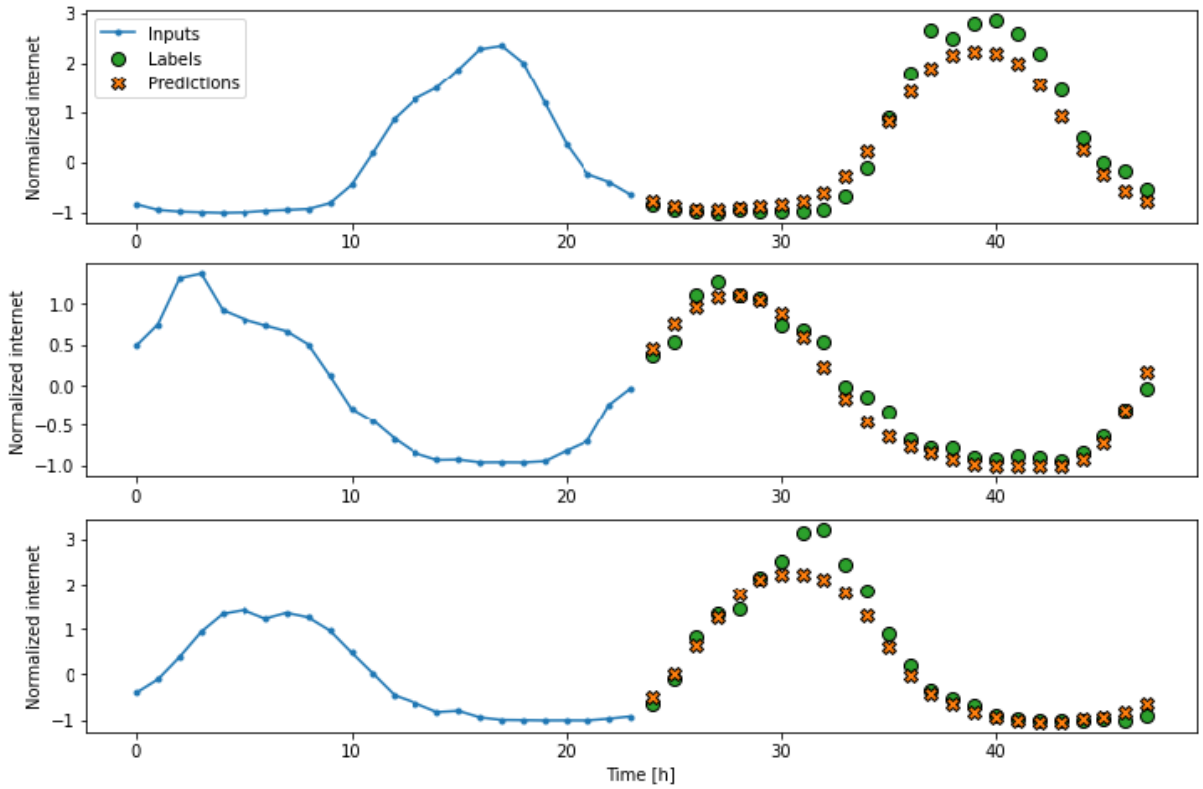


Figure 23 - Prediction of the ARLSTM model on "grid 5161" normalized internet traffic.

Model	Averaged MAE
Baseline Model	0.39
FCSN	**0.29**
1D-CNN	**0.29**
SS-LSTM	0.32
AR-LSTM	0.32

Table 7 - Average MAE for all features in all grids

The predictive performance of temporal models including temporal baseline, FCSN, 1DCNN, and SSLSTM was investigated. The ARLSTM model is not used for the prediction of certain types of mobile traffic. The ARLSTM model fails in mobile traffic forecasting when the number of features decreases as the model output is fed back at each phase to be used as input for the forecast of the next time track. The temporal models of cellular traffic forecasts for certain types of cellular traffic including internet, sms, and calls are summarized in Table 8. Based on the table, 1DCNN outperforms other models with the smallest MAE and RMSE. But for sms prediction, the three models have comparable performance. In terms of call traffic prediction, the work of the SSLSTM model is marginally improved than other temporal models. However, the 1DCNN model has a smaller model size with less complexity and is more adequate to use for edge computing deployments.

p-ISSN : 2087-0868 e-ISSN : 2598-9707

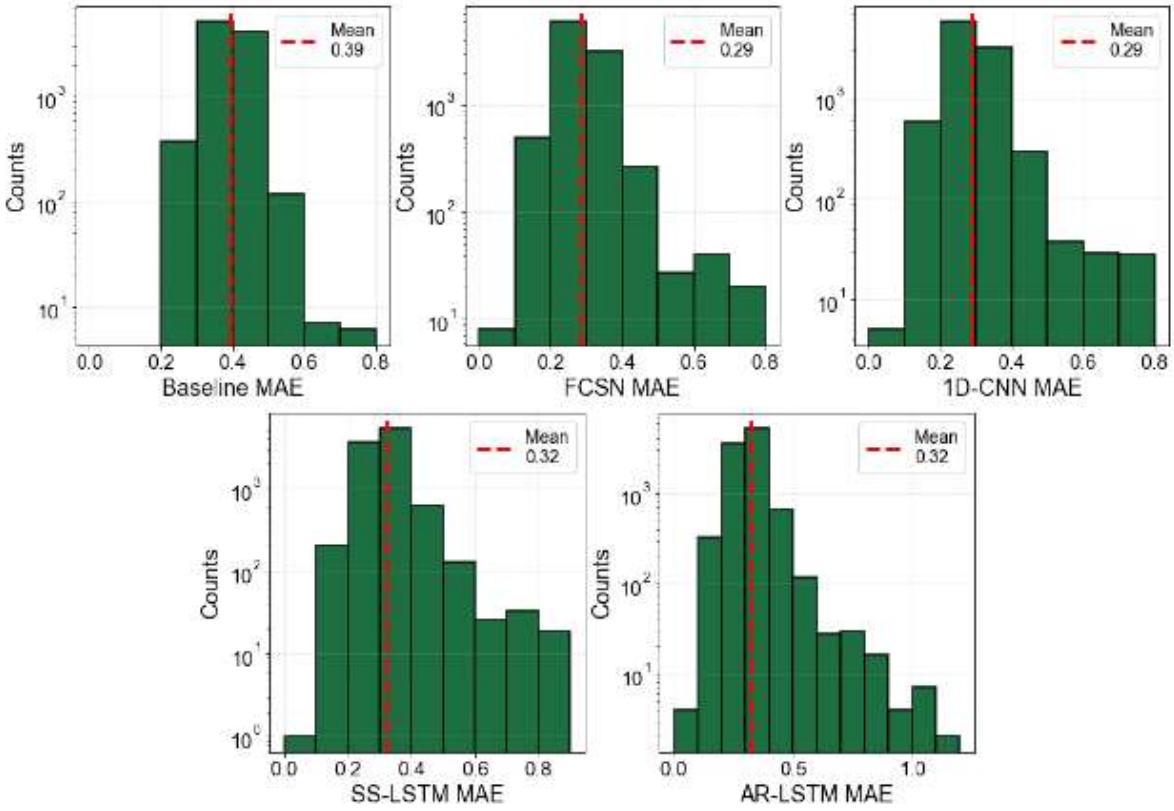


Figure 24 - MAE distribution of the proposed model across all grids.

Traffic	Model	MAE	RMSE
Internet	Temporal baseline	153.72	205.95
	FCSN	117.39	152.66
	1D-CNN	**113.54**	**147.43**
	SS-LSTM	124.32	160.93
Sms	Temporal baseline	30.07	45.56
	FCSN	17.96	**32.30**
	1D-CNN	17.60	32.62
	SS-LSTM	**17.10**	32.66
Call	Temporal baseline	27.02	36.62
	FCSN	14.92	22.42
	1D-CNN	14.77	22.26
	SS-LSTM	**13.08**	**21.25**

Table 8 - Predictions of mobile traffic temporal models

In this spatial distribution, high MAE is shown in red and low MAE is depicted in blue. Therefore, a darker blue color represents less error. Finally, to get a comprehensive assessment of the proposed forecasting model. As an outlook, the base model was the fastest with a run-time of 0.31 seconds, and the ARLSTM had a great hanging time of 24.31s. the FCSN and 1DCNN have been confirmed in table 9, which have significant differences in execution time compared to the LSTM approach.

p-ISSN : 2087-0868 e-ISSN : 2598-9707

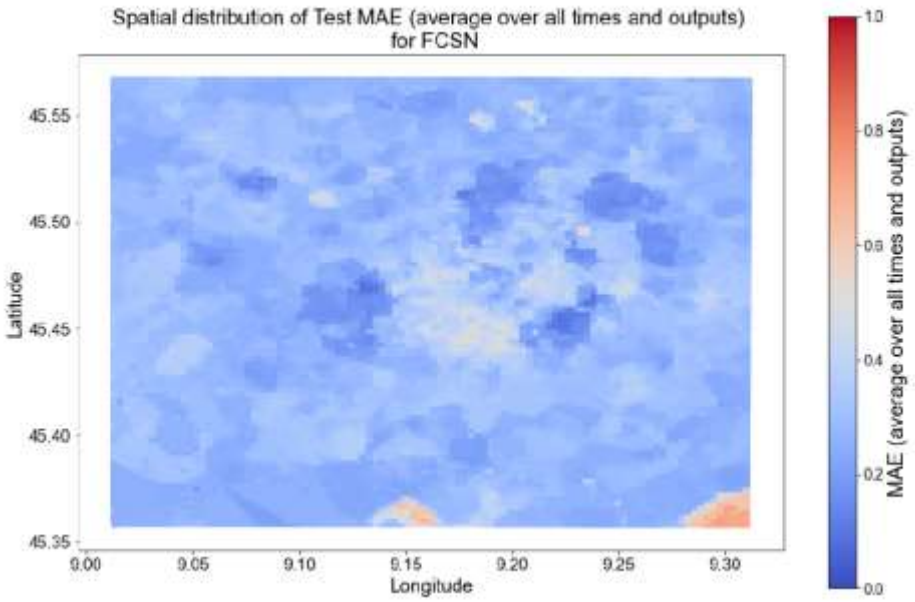


Figure 25 - dimensional delivery of MAE for prediction models of FCSN.

Model	Execution time (seconds)
Baseline Model	0.31
FCSN	6.98
1D-CNN	6.37
SS-LSTM	21.38
AR-LSTM	24.31

Table 9 - Execution time of the proposed prediction model

Spatiotemporal mobile traffic prediction

For spatiotemporal mobile traffic prediction, 70% of the dataset (1040 CDR) was used for the instruction set 298 records for the verification set and 149 records for the test set. In addition, Min-Max scaling is applied to the data set to scale the input range between the ranges in [0, 1]. The 2D-ConvLSTM model is trained considering 500 epochs with early termination monitoring for loss of validation. The optimizer Adam and MAE as a loss function is used during the training process. Internet, sms, and call predictions using the 2D-ConvLSTM model in all grids for 61-time steps are depicted in Figure 26, Figure 30, and Figure 28, respectively. Internet cellular traffic, sms, and calls' performance for the spatiotemporal baseline and 2D-ConvLSTM are presented in Table 10. In addition, the proposed model execution time on all types of mobile traffic including internet, sms, and calls is illustrated in Table 11.

p-ISSN : 2087-0868 e-ISSN : 2598-9707

Table 10 - 2D-ConvLSTM Hyperparameters.

Criteria / Model	Loss		Optimizer			Learning rate		
	MAE	MSE	Adam	Adamax	SGD	0.1	0.01	0.001
2D-ConvLSTM	✓	x	✓	x	x	x	x	✓
Model structure	• Considering RELU, Tanh, and **sigmoid** activation functions. • Considering various numbers of hidden layers starting from 1, 2, 3, 4, and **5**. • Considering different kernel sizes including **(10 × 10)**, **(8 × 8)** , (6 × 6), **(5 × 5)**, **(3 × 3)**. • Considering different filter sizes containing **10**, **8**, 6, **5**, and **3**.							

Table 11 - Different types of mobile traffic performance

Traffic	Model	MAE	RMSE
Internet	Spatiotemporal baseline	102.12	142.41
	2D-ConvLSTM	**52.73**	**75.73**
Sms	Spatiotemporal baseline	24.03	36.04
	2D-ConvLSTM	**14.42**	**26.60**
Call	Spatiotemporal baseline	15.23	22.06
	2D-ConvLSTM	**8.98**	**15.02**

Table 12 - The proposed model execution time on other varieties of traffic

Traffic	Model	Execution time (seconds)
Internet	Spatiotemporal baseline	10.15
	2D-ConvLSTM	4419.16
Sms	Spatiotemporal baseline	6.89
	2D-ConvLSTM	2821.77
Call	Spatiotemporal baseline	12.87
	2D-ConvLSTM	3406.08

p-ISSN : 2087-0868 e-ISSN : 2598-9707

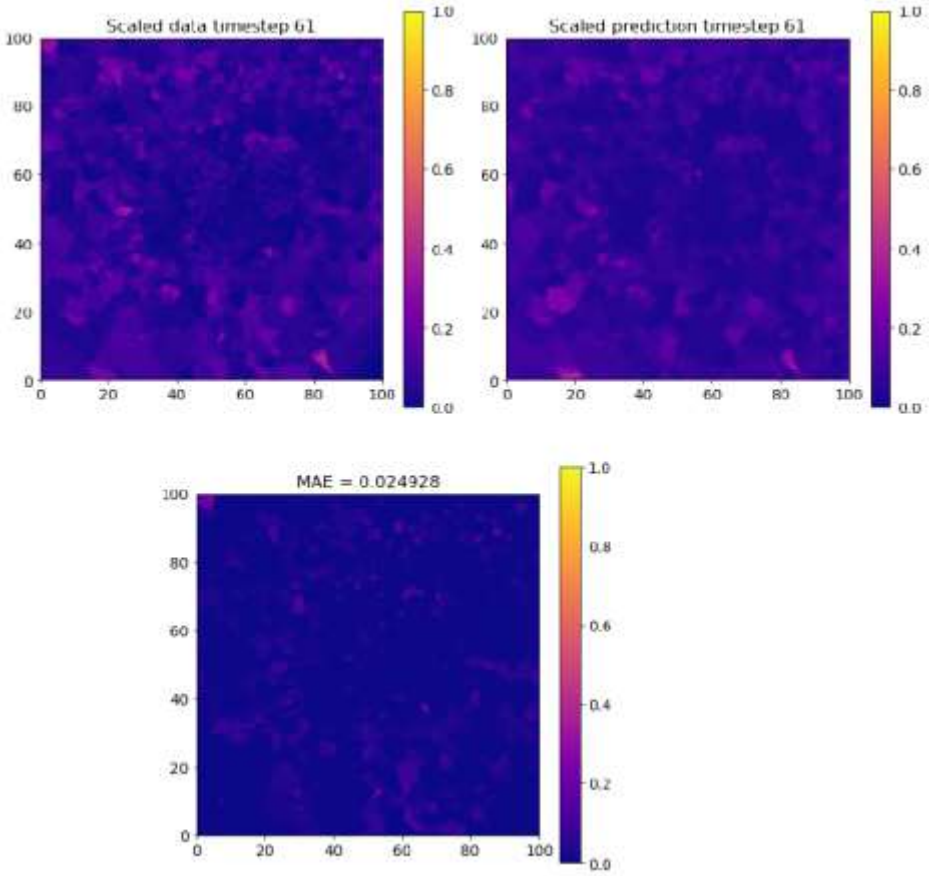


Figure 26 - Internet prediction performance of 2D-ConvLSTM at 61-time steps.

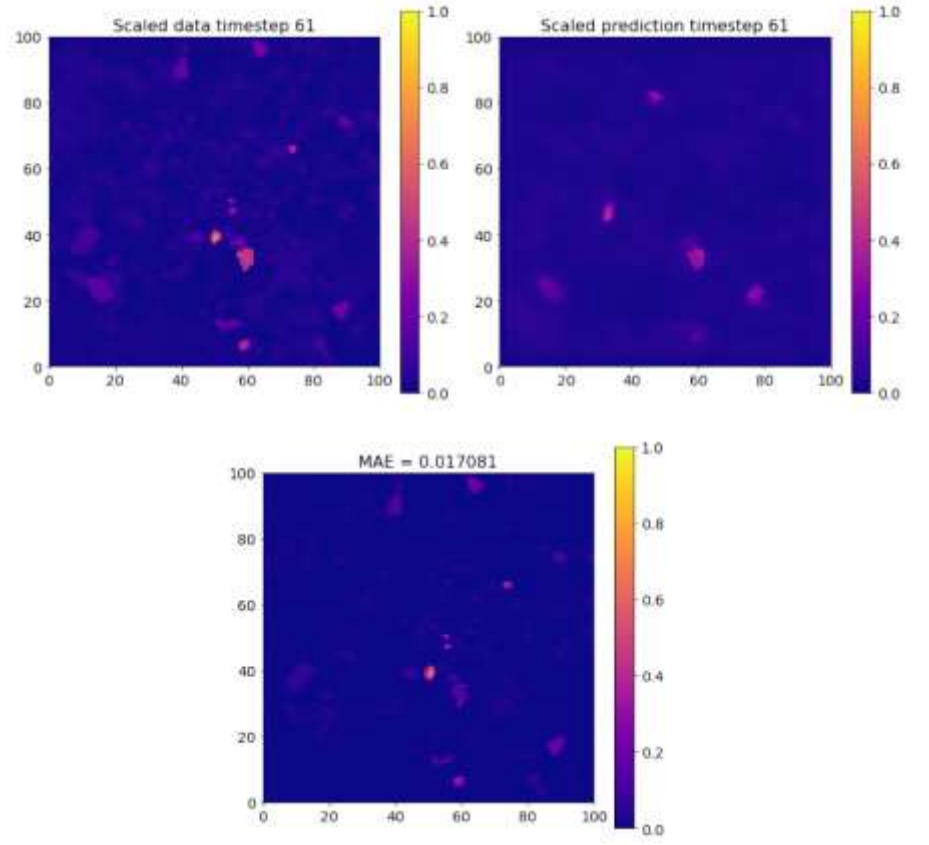
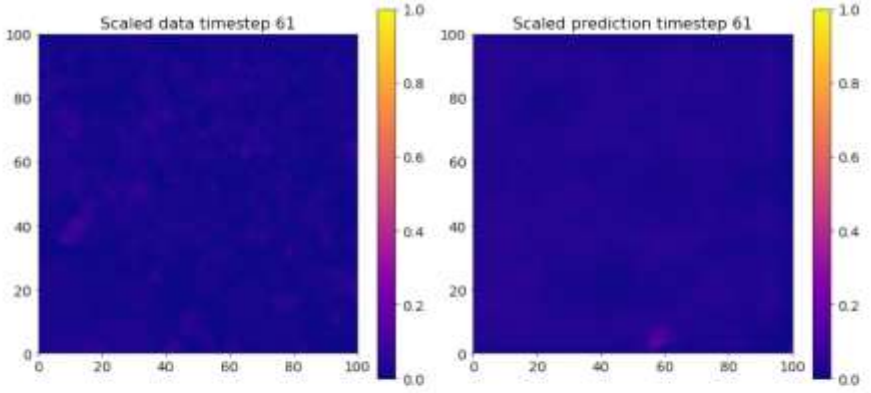


Figure 27 - 2D-ConvLSTM sms prediction performance at 61-time steps.



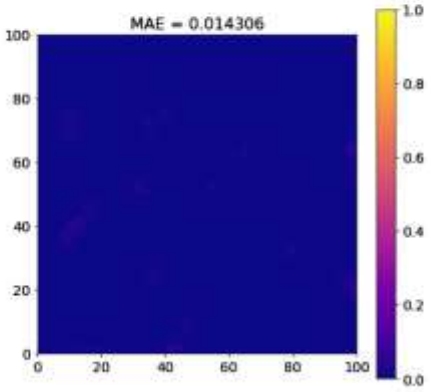


Figure 28 - 2D-ConvLSTM call prediction performance at 61-time steps.

Another factor that contributes to the success of the proposed model may be implementing a new measure named "count" along with "sms", "calls", and "internet" as model inputs. Utilizing counts helps predict different types of mobile traffic more accurately because it shows the total of files at a given pace for a given network's generative force.

CONCLUSION

The main focus Of this study is the prediction of mobile traffic 24 hours ahead by utilizing temporal and spatiotemporal approaches and the predictive work of a lot of neural network models is assessed. The special temporal framework consists of FCSN, 1DCNN, SSLSTM, and ARLSTM. For the spatiotemporal framework, the 2D-ConvLSTM model is proposed to forecast cellular traffic for 24 hours next. Among the temporal models, FCSN and 1DCNN have proportionate work with the smallest MAE. Nonetheless, 1DCNN is a smaller network with less number of parameters. In addition, the proposed 1DCNN has lower complication and exhibits a lower hit time for forecasting the traffic in 24-hour next. In terms of spatiotemporal prediction, the proposed 2DConvLSTM model shows better performance for predicting all three types of mobile traffic, including internet, sms, and calls compared to

p-ISSN : 2087-0868 e-ISSN : 2598-9707

temporal analysis showing the effectiveness of incorporating data dependencies in traffic prediction. It is important to recognize the ability demands of every slice and how these requirements vary supplementary. Accurate model predictions help avoid provisioning shortfalls that lead to poor network slice performance and poor QoS for users. In addition, over-provisioning can result in costs for infrastructure providers. Therefore, since dynamic adjustment of resource allocation to network slices in a 5G network is required, predicting the traffic profile of each slice is very important.

The results of this study indicate that “FCSN” and “1DCNN” it has proportionate conduct on time series models. Nonetheless, the SSLSTM model outperforms other temporal models in terms of MAE and RMSE metrics in terms of call traffic forecasting. Regarding sms traffic, SSLSTM has the lowest MAE while FCSN is better in terms of RMSE. Moreover, 1DCNN outperforms other temporal models for forecasting internet traffic in terms of MAE and RMSE. In addition, the expected 1DCNN is less complex and has a quick execution time to predict the traffic for 24 hours next. Therefore, it is expected that network optimization and more effective resource allocation can be carried out by predicting cellular traffic through the proposed 2D-ConvLSTM model.

Suggestions For Future Research

In future work, the prediction performance can be improved and the deployment of the 2D-ConvLSTM model in 5G networks can be optimized automatically. The costs incurred by MNO to allocate resources to each slice before and after using the 2D-ConvLSTM model for mobile traffic forecasting can also be calculated and compared. Future research may utilize other data sets to measure the model generalizability of the proposed spatiotemporal 2D-ConvLSTM model. A potential next step is to use more historical data that can be used to estimate different types of mobile traffic. Mobile traffic prediction on holidays will help MNO in terms of appropriate resource allocation before different holidays every year. In addition, the predictive ability of the proposed model over a longer time frame will be tested in future work. Additionally, Pruning and optimizing models for edge applications can also be explored. Statistical analysis to identify extreme values and consider how the proposed model can handle outliers and consider network performance influences such as latency in mobile traffic prediction.

BIBLIOGRAPHY

Afolabi, I., Ksentini, A., Bagaa, M., Taleb, T., Corici, M., & Nakao, A. (2017). Towards 5G Network Slicing over Multiple-Domains. https://publica.fraunhofer.de/handle/publica/253790

Ali Imran, Ahmed Zoha, and Adnan Abu-Dayya. Challenges in 5g: how to empower son with big data for enabling 5g. IEEE networks, 28(6):27–33, 2014.

Chunxiao Jiang, Haijun Zhang, Yong Ren, Zhu Han, Kwang-Cheng Chen, and Lajos Hanzo. Machine learning paradigms for next-generation wireless networks. IEEE Wireless Communications, 24(2):98–105, 2016.

Dario Bega, Marco Gramaglia, Marco Fiore, Albert Banchs, and Xavier Costa-Perez. Deepcog: Cognitive network management in sliced 5g networks with deep learning. In IEEE INFOCOM 2019-IEEE Conference on Computer Communications, pages 280– 288. IEEE, 2019.

Debasis Bandyopadhyay and Jaydip Sen. Internet of things: Applications and challenges in technology and standardization. Wireless personal communications, 58(1):49–69,2011.

Duong D. Nguyen, Hung X. Nguyen, and Langford B. White. Reinforcement learning with network-assisted feedback for heterogeneous rat selection. IEEE Transactions on Wireless Communications, 16(9):6062–6076, 2017.

Fabio Giust, Luca Cominardi, and Carlos J Bernardos. Distributed mobility management for future 5g networks: overview and analysis of existing approaches. IEEE Communications Magazine, 53(1):142–149, 2015.

Fengli Xu, Yuyun Lin, Jiaxin Huang, Di Wu, Hongzhi Shi, Jeungeun Song, and Yong Li. Big data driven mobile traffic understanding and forecasting: A time series approach. IEEE transactions on services computing, 9(5):796–805, 2016.

Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. science, 313(5786):504–507, 2006.

Giuseppe A Carella, Michael Pauls, Thomas Magedanz, Marco Cilloni, Paolo Bellavista, and Luca Foschini. Prototyping nfv-based multi-access edge computing in 5g ready networks with open baton. In 2017 IEEE Conference on Network Softwareization (Net-Soft), pages 1–4. IEEE, 2017.

Jing Wang, Jian Tang, Zhiyuan Xu, Yanzhi Wang, Guoliang Xue, Xing Zhang, and Dejun Yang. Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach. In IEEE INFOCOM 2017-IEEE Conference on Computer Communications, pages 1–9. IEEE, 2017.

Jong-Hyouk Lee and Hyoungshick Kim. Security and privacy challenges in the internet of things [security and privacy matters]. IEEE Consumer Electronics Magazine, 6(3):134–136, 2017.

p-ISSN : 2087-0868 e-ISSN : 2598-9707

Kan Zheng, Zhe Yang, Kuan Zhang, Periklis Chatzimisios, Kan Yang, and Wei Xiang. Big data-driven optimization for mobile networks toward 5g. IEEE networks, 30(1):44–51, 2016.

Mamta Agiwal, Abhishek Roy, and Navrati Saxena. Next generation 5g wireless networks: A comprehensive survey. IEEE Communications Surveys & Tutorials, 18(3):1617–1655, 2016.

Mohammad Abu Alsheikh, Dusit Niyato, Shaowei Lin, Hwee-Pink Tan, and Zhu Han. Mobile big data analytics using deep learning and apache spark. IEEE networks, 30(3):22–29, 2016.

Osianoh Glenn Aliu, Ali Imran, Muhammad Ali Imran, and Barry Evans. A survey of self organization in future cellular networks. IEEE Communications Surveys & Tutorials, 15(1):336–361, 2012.

P´eter Szil´agyi and Szabolcs Nov´aczki. An automatic detection and diagnosis framework for mobile communication systems. IEEE transactions on Network and Service Management, 9(2):184–197, 2012.

Parthasarathy Guturu. Explosive wireless consumer demand for network bandwidth fifth generation and beyond [future directions]. IEEE consumer electronics magazine, 6(2):27–31, 2017.

Prakash Suthar, Vivek Agarwal, Rajaneesh Sudhakar Shetty, and Anil Jangam. Migration and interworking between 4g and 5g. In 2020 IEEE 3rd 5G World Forum (5GWF), pages 401–406, 2020.

Ravishankar Ravindran, Asit Chakraborti, Syed Obaid Amin, Aytac Azgin and Guoqiang Wang 5G-ICN: Delivering ICN Services over 5G using Network Slicing Huawei Research Center, Santa Clara, CA, USA.

Sepp Hochreiter and J¨urgen Schmidhuber. Long short-term memory. neural computing , 9(8):1735–1780, 1997

Tiago Prado Oliveira, Jamil Salem Barbar, and Alexsandro Santos Soares. Computer network traffic prediction: a comparison between traditional and deep learning neural networks. International Journal of Big Data Intelligence, 3(1):28–37, 2016.

Tuyen X. Tran, Abolfazl Hajisami, Parul Pandey, and Dario Pompili. Collaborative mobile edge computing in 5g networks: New paradigms, scenarios, and challenges. IEEE Communications Magazine, 55(4):54–61, 2017.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. nature, 518(7540):529–533, 2015.

Wei Jiang, Mathias Strufe, and Hans D Schotten. Experimental results for artificial intelligence-based self-organized 5g networks. In 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), pages 1–6. IEEE, 2017.

Xiaofei Wang, Xiuhua Li, and Victor CM Leung. Artificial intelligence-based techniques for emerging heterogeneous networks: State of the arts, opportunities, and challenges. IEEE Access, 3:1379–1391, 2015.

Xin Li, Mohammed Samaka, H Anthony Chan, Deval Bhamare, Lav Gupta, Chengcheng Guo, and Raj Jain. Network slicing for 5g: Challenges and opportunities. IEEE Internet Computing, 21(5):20–27, 2017.

Young-il Choi and Noik Park. Slice architecture for 5g core network. In 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN), pages 571–575, 2017.

Yun Chao Hu, Milan Patel, Dario Sabella, Nurit Sprecher, and Valerie Young. Mobile edge computing—a key technology towards 5g. ETSI white paper, 11(11):1–16, 2015.