Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 7 No. x (2023) 1332 - 1340 ISSN Media Electronic: 2580-0760

Linear and Non-Linear Spatio-Temporal Input Selection In Wireless Traffic Networks Prediction using Recurrent Neural Networks

Ahmad Saikhu[1], Agung Teguh Setyadi[2]*, Victor Hariadi[3]
[1,3]Teknik Informatika, Fakultas Teknologi Elektro dan Informatika Cerdas, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2]Rekayasa Perangkat Lunak, Fakultas Sains dan Teknologi, Universitas Anwar Medika, Sidoarjo, Indonesia
[1]saikhu@if.its.ac.id, [2]agung.teguh.setyadi@gmail.com, [3]victor@its.ac.id

Abstract

For the optimization of computer networks with high bandwidth requirements, wireless network traffic prediction is necessary. Its goal is to reduce maintenance costs and enhance internet services. Feature selection is a major issue in the Multivariate Time Series (MTS) Spatio-temporal modeling. Another problem is the dependency between input features, time-lags, and spatial factor so that an appropriate model is needed. This study aims to provide solutions to two problems. The first is to improve a feature extraction and selection process in Spatio-temporal MTS data for relevant features using Detrended Partial Cross-Correlation Analysis (DPPCA) and non-redundant features associated with linear using Pearson's Correlation (PC) filters and non-linear associations using Symmetrical Uncertainty (SU) and combination of both PCSUF. The second is to develop a Spatio-temporal framework model using Recurrent Neural Networks (RNN) to get a better performance than traditional model. These methods are combined and tested using the dataset of cellular networks with one-hour intervals during November in three locations. Testing the effectiveness of the feature selection technique showed that 27.6% of the total extracted features. The forecasting model with the DPCCA–SU-RNN combination method gets the best performance by having RMSE = 380.7, R^2 = 97%, and MAPE = 10%.

Keywords: *wireless traffic; linear and non-linear; spatio-temporal; recurrent neural network*

1. Introduction

In human life, we usually find time-series data in various sectors. This data can be found, among others, in weather/climate observations, production monitoring, economic/financial activities, medical records/diagnoses, earthquake observations, sensor data, and computer network traffic data. This data is available in abundance and requires analysis/ modeling tools to benefit from recording the data, including decision-making. Time series data can be in the form of a single variable, which is called univariate. The other is a recording that includes many variables, called MTS.

Wireless networks are changing the system of the networking world nowadays. The rapid advancement of mobile networks and the need for open internet access suggest that most individuals are becoming used to life online. However, when a wired network is inaccessible due to unexpected wireless network activity factors, wireless network services are also necessary and impact service quality guarantees.

Network service activity forecasting is useful for designing, operating, controlling, and improvement of communication networks. Optimization is achieved by preventing congestion, detecting network traffic instabilities, monitoring response, and allocating system resources. Its purpose is to decrease costs and enhance bandwidth services due to persistent high system demands [1].

The number of researches studies show that in standard theory, system network activity characteristics extend beyond the behavior model paradigm of Poisson and Markov [2]. The absence of theory and research to model multivariate, Spatio-temporal, and non-linear system network behavior makes this research continue to grow. Researchers have the objective to define real data traffic to achieve a proper and capable wireless network traffic model. Its advantages include the estimation of network traffic levels, which is useful under similar conditions for planning.

Varieties, the MTS description system, categorization, pattern detection, identification of anomalies, and

Accepted: 23-07-2023 | Received in revised: 25-11-2023 | Published: 12-12-2023

forecasts are the criteria' MTS analysis tool. The MTS model predictions are significant for MTS modeling and many researchers in various application areas and the last couple of years. Method MTS predictions are important MTS modeling and many researchers in the previous decade in different application fields. Some recent studies, namely Spatio-temporal, have implemented different approaches to modeling the Multivariate time series data estimator. Faghmous and Kumar [3] suggested that Spatio-Temporal Data Mining is used to improve precision, scalability, and easy understanding of the multivariate periodic series problem.

The statistics time-series model, such as Autoregressive, Artificial Neural Network (ANN), and wavelet-based methods, is used to support research on chaotic network behavior predictive models [4]. The Multivariate Statistical Network Monitoring (MSNM) method is additionally used for detection intrusions and anomalies [5], the Deep Belief Networks (DBN) method for predicting network traffic data [6], and the Fuzzy Machine Neural Networks (FMNN) method for regulation of network traffic. Alternatively, use a combination model of Stationary Wavelet Transform (SWT), Quantum Genetic Algorithm (QGA), and Back-propagation (BP) to achieve higher service levels with a lot of efficient prediction model [1].

In consumer traffic prediction research on cellular-based machine-learning, an MTS with a Spatio-temporal approach was developed to predict traffic data of wireless network [7]. Other research is multi-layer prediction models Feed Forward Neural Network (FFNN) for multi-channel of wireless network traffic case [8]. In big data cellular networks, a wireless network prediction model approach considers that complex, non-stationary, and non-linear data activity characteristics have also been performed using Quantification Analysis (QA) and Information Entropy (IE) and a prediction model with a Spatio-temporal in-depth learning approach.

Several studies use deep learning to predict cellular network dataset Spatio-temporal data-deep learning system LA-Resnet based on a mechanism of attention proposed by [8]. There are three phases of the LA-Resnet framework: residual network, Recurrent Neural Networks (RNN), and an attention mechanism. To extract the spatial features of data, researchers commonly use the residual network stage. The RNN stage and the attention mechanism, on the other hand, are used to collect data on temporal dependencies.

Another study used Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) models to forecast mobile internet traffic in each Milan region or grid that has been clustered based on internet activity using the K-means method [9]. The study [10] proposed deep regression that can extract multiscale spatiotemporal correlation. The Multi-Channel Spatial-Temporal Framework that consists of 3 channels Convolution Neural Networks (CNN) and LSTM, was proposed by [9]. The three channels CNN in that study was used to overcome the spatial dependencies of cellular data. The study [11] proposed a Hybrid Spatio-Temporal Network (HSTNet). HSTNet is a CNN-based deep learning framework using a deformable convolution unit to increase the ability to extract complicated spatial characteristics. From the experiments they did, each of the methods they proposed could increase the prediction accuracy.

To obtain a forecasting model with an efficient and high accuracy process depends on the MTS data application's real and the appropriate method's choice. Machine learning-based MTS is still better approach in some studies than others [12]. It is not easy to do MTS modeling since it requires several input variables with tens or even hundreds of characteristics. For the answer variable, not all input features are essential. Not all of the input features or predictors are relevant and significant to the response variable. For the output variable, not all input features are essential. A method is needed to pick the appropriate set of features in the prediction step and dismiss some other features [13]. A case occurs from the applicable feature set where the input feature is redundant such that the forecasting mechanism is unreliable, and it can reduce the accuracy of the forecasting. In the course of selecting input features, this question must also be resolved.

Therefore, this study provides several solution to overcome these problems. The first is to improve a feature extraction and selection process in Spatio-temporal MTS data for relevant features using Detrended Partial Cross-Correlation Analysis (DPPCA) and non-redundant features associated with linear using Pearson's Correlation (PC) filters and non-linear associations using Symmetrical Uncertainty (SU) and combination of both PCSUF. The second is to develop a Spatio-temporal framework model using Recurrent Neural Networks (RNN) to get a better performance than traditional model. The predicting model is RNN, where an RNN appropriate for data of Spatio-temporal MTS is developed. Performance testing of the proposed prediction model framework will be carried out by comparing predicted data with actual data.

2. Research Methods

The dataset and some methods are discussed in this section. These methods included the predictor's extraction using DPCCA from a data of Spatio-temporal MTS, the collection of appropriate features, and the elimination of redundant predictor using PCSUF, modelling time series data using RNN model, measurement training and forecasting performance.

2.1 Case Study

This study uses a dataset obtained from the Big Data Challenge Telecom Italia [14]. The dataset consists of cellular networks and other urban data covering the Milan area. However, this study only uses a cellular network dataset. The cellular network dataset in the City of Milan has a period from 00:00 on November 1st, 2013, to 23:59 on January 1st, 2014, with 10-minute intervals. The cellular network dataset in Milan has divided into 100 × 100 grids, as shown in Fig. 1. Each grid has a size of 235 x 235 square meters. Grids in the dataset were estimated using the World Geodetic System 1984 (WGS84) standard (EPSG: 4326) and expressed in GeoJson format. Each grid is an aggregation of existing data from the area of the grid. The dataset information shown in Table 1.

Table 1.Dataset Details

Data	Details
Square id	the identification string of each Grid in Milan
Time interval	the time intervals on the dataset with 10-minute intervals
Country code	the country code of the phone number
SMS-in	the number of incoming message activities on a particular grid (X1)
SMS-out	the number of incoming message activities on a particular grid (X2)
Call-in	the number of incoming call activities on a particular grid (X3)
Call-out	the number of incoming call activities on a particular grid (X4)
Internet traffic	the amount of internet usage by mobile users on a particular grid (Y) is generated if the connection takes more than 15 minutes, or more than 5 MB is transferred during the session

This study only used the dataset of cellular networks during November. At the outset, the dataset is cleaned and pre-processed for missing data. Then, remove the country code feature because it does not affect the amount of internet traffic data. The dataset is aggregated into 1-hour intervals by sum each data at the same hour because the data too sparse. This study selects three grids based on the correlation of internet traffic data shown in Table 2. The distance between three grids shown in Table 3.

Table 2.Grids based on the correlation

Grid_ID	4259	5060	5160
4259	1	0.77	0.80
5060		1	0.95
5160			1

Table 3.Grids based on the distance

Grid_ID	4259	5060	5160
4259	0	2115 m	2350 m
5060		0	235 m
5160			0

Figure 1 shows 100 hourly internet traffic distribution start from Monday on November 4th, 2013 for three grids. Figure 2 show 3d heat map for 1 month of internet traffic data for each grids.

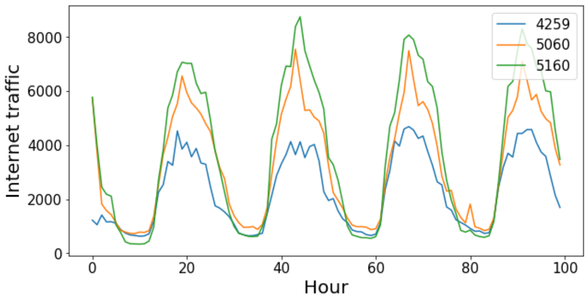

Figure 1.The hourly distribution of internet traffic data at 3 Grids

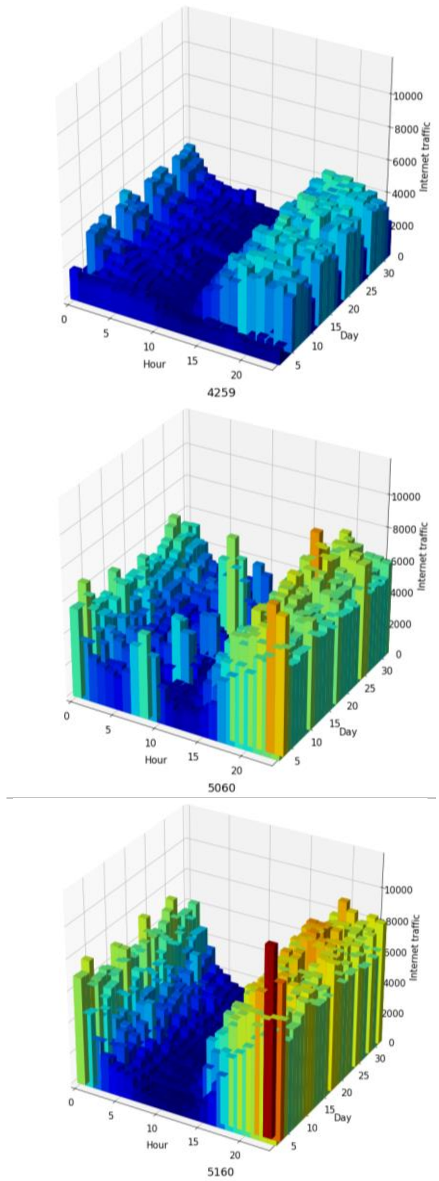

Figure 2.The daily distribution of internet traffic data at 3 Grids

Based on Figure 1 and Figure 2, the distribution of internet data traffic on the three grids has a similar pattern, which indicates the existence of traffic dependencies between the grids. This is reinforced by the high correlation value between the three grids in Table 2, which shows that spatio-temporal-based

DOI: https://doi.org/10.29207/resti.v7i6.5296

forecasting modeling for wireless traffic networks is more suitable.

2.2 Spatio-temporal dataset

A statistic could be a time-based series of knowledge items connected to every different, written as: $X_N(t)$; [l = 1, 2, ..., N, t = 1, 2, …, T]. If we add location factor and N > 2, it is recognized as a spatio-temporal of multivariate time-series. Spatio-temporal internet-traffic data Y_{tl}, at time t, t = 1, 2, …, T, at grid-location l, l = 1, 2, …, L can be formulated as Formula 1 [15].

$$Y_{tl} = \theta_t + m_l + \varepsilon_{tl} \tag{1}$$

Time series contain stochastic process data, which cannot be modeled by linear approximation. Non-linear properties occur in stationary and non-stationary time station data [16]. If any amendment within the price of x tends to be followed by y with a relentless quantitative relation, two variables x and y have a linear connection [17], [18]. The result will be a straight-line pattern if the linear relationship x and y area unit aforethought in a very scatter diagram, supported the white-test for non-linearity, and the relationships between input options and response options non-linear [19]. This situation is named 'relevant non-linear'.

2.3 DPCCA

DPCCA eliminates the influence of other time series operating as joint powers and uncovers the intrinsic power-law cross-correlations between two concurrently reported time series in the existence of non-stationary data [20]. DPCCA has the benefit of disclosing 'intrinsic' relationships between two different time series of interest, with the exclusion of possible influences from other unconsidered signals [21]. DPCCA guarantees that the results achieved are not influenced by patterns, including linear, exponential, and also trends in higher-order and trend periodicals, by detrending local trends. More implicit correlation information than other research approaches can be uncovered by Detrended Cross Correlation Analysis (DCCA), leading to its adoption and implementation in different fields [22].

The DPCCA value (ρ) is determined as the detrended covariance of two datasets C_{x_{ij},y_j} over two integrated series detrended variance $C_{x_{ij},y_{ij}}$ and C_{x_j,y_j} [20] with size of box s, as Formula 2.

$$\rho DPCCA\left(x_{ij}, y_j, s\right) = \frac{-C_{x_{ij},y_j}(s)}{\sqrt{C_{x_{ij},x_{ij}}(s), C_{y_i,y_i}(s)}} \tag{2}$$

In this modeling, the predictor variable is extracted from each location and the effect of time-lag into a three-dimensional array feature $N \times T \times L$. Furthermore, the DPCCA value is calculated from the predictor feature x_{ij} with the response variable. The candidate feature predictors are those with significant DPCCA scores.

2.4 Pearson's Correlation and Symmetrical Uncertainty Filter (PCSUF)

Three mutually exclusive types can be categorized into the interactions between output responses and predictor: irrelevant behaviors, significant yet redundant behaviors, and relevant characteristics [23]. The first feature group are not classified as a variable predictor, so it does not need to be chosen. The second type, i.e., essential and not redundant features, should be selectively chosen because it is necessary to verify whether it is relevant. As for the last group, since it is a valid independent predictor for other features, a function that is an absolute must be a predictor, as shown in Figure 3.

The purpose of selecting the predictor feature candidates using the PC Filter, SU Filter, and a combination of both PCSUF is to eliminate the irrelevant predictor feature, reduce redundancy so that the number of predictor features is minimal and the value of relevance to the response variable is maximized.

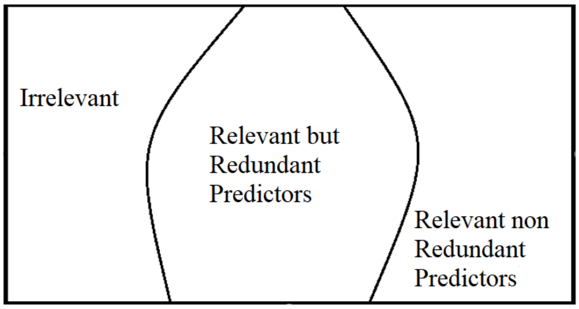

Figure 3.Irrelevant, relevant, and redundant features

To use PC as described in Formula 3, we test significant and unnecessary entities in a subset as a linear relation in which ρ_{xy} is a correlation for both x variable and y variable. If its coefficient is larger than a critical correlation table, ρ_{xy} is declared significant.

$$\rho_{xy} = \frac{\sum_i (x_i - \bar{x})(y_i - y)}{\sqrt{\sum_i (x_i - \bar{x})}\sqrt{\sum_i (y_i - \bar{y})}} \tag{3}$$

To evaluate a non-linear relationship between predictor variables and the response variable, SU is used. SU is developed on the basis of entropy and data gain. If we have got the x variable, therefore the entropy is described as explained in Formula 4 and 5 [23], where the prior probability for all X value is $P(xi)$, and the posterior probability of X is $P(xi|yi)$ due to the importance of Y.

$$H(Y) = \sum_i P(x_i) \log(P(x_i)) \tag{4}$$

$$H(Y) = -\sum_j P(x_i) \sum_i P(y_i) \log(P(y_i)) \tag{5}$$

The entropy of variable X is defined as expressed in Formula 6 after analyzing the values of another

DOI: https://doi.org/10.29207/resti.v7i6.5296

variable, Y. Additional X information provided by Y provides the sum by which the X entropy decreases.

$$IG(Y) = H(X) - H(Y) \quad (6)$$

Nevertheless, information gain is biased in favor of features with more values. Also, to ensure that they are identical and have the same effect, feature values must be standardized. We select SU Formula 7, which is appropriate for discrete data. We must do the discretization process if the data is continuous. If its value is greater than the critical value of the table of Chi-Square with an absolute DOF and SL, SU(X, Y) is considered necessary.

$$SU(X,Y) = 2\left[\frac{IG(Y)}{H(X)+H(Y)}\right] \quad (7)$$

To know whether a predictor group is irrelevant, relevant, or redundant, we use the measurement of non-linear correlation, SU to Formula 7. The use of the critical value $X^2\alpha$. table is to test SU's results (X, Y).

We used merit-value in this research to assess the performance of PC, SU, and PCSU feature selection. Numerous previous researches have used merit-value to measure the optimal performance of feature selection, such as the Maximum Relevancy and Minimum Redundancy-Hesitant Fuzzy Sets (MRMR-HFS) feature selection algorithm [24], recognition of human motion sequence [25], epileptic disease classification [26] and forest-fire modeling [27] to achieve the high performance of the built models.

Table 4 explains the process of selecting the predictor feature and removing redundancy using the PC filter. Table 5 explains the predictor feature selection process using SU Filter. While the selection process using the PCSUF is a combination of the two methods by looking for a combination of α and β values that causes the maximum merit-value $\pi_{X_{ij},Y}$ with optimal value.

2.5 Time-series forecasting using Spatio-temporal RNN

Adaptive control problems, machine detection, speech recognition, and time series forecasting were solved with RNN model [28]. The architecture of an RNN is identical to that of a standard multilayer perceptron, with the only difference being that time delays may be correlated with a connection between hidden units. Models can store information from the past with this network architecture, thus finding the data relationship between events temporarily distant from each other [29].

Table 4. Pseudocode of input selection using PC method

```
Pseudo-Code of PC-Filter
Input:
S(X1, X2, ….., Xn,) //  spatio-temporal predictor set with three dimension : N features x T time-lags x L locations;
Y // Response Variable;
α         // a predefined correlation threshold
Output: S1list', α', πij
            Begin
               For i = 1 to n do begin
                  Calculate ρi,Y for Xi;
                  If (ρi,Y >  correlation critical table)
                        add Xi to S1list;
            End
            Order S1'list in descending
            Xj= putFirstElement at (S1'list)
            do begin
                 Xi=putNextElement(S1'list, Xj)
                 if (Xi <> Null)
                 do begin
                      If (ρi,Y > α)
                      remove Xi from S1'list
                      Xi=putNextElement(S1'list, Xi)
                  End until (Xi==Null)
                  Xj=putNextElement(S1'list, Xj)
            End until (Xj==Null)
            Count πXij,Y
            α '= Update(α)
            repeat (8) – (19) until optimal Count πXij,Y
            Sbest= S1'list with optimal Count πXij,Y value
```

Table 5. Pseudocode of input selection using SU Method

```
Pseudo-Code of SU-Filter
Input:
S(X1, X2, ….., Xn,) //  spatio-temporal predictor set with three dimension : N features x T time-lags x L locations result of DPCCA;
n // number of predictor features
Y // Response Variable;
β // a threshold of SU
Output: S2list', β', πij
            Begin
               For i = 1 to n do begin
                  Calculate SUi,Y for Xi;
                  If (SUI, Y > value of critical table χ2)
                        add Xi to S1list;
               End
               Order S2'list in descending
            Xj= putFirstElement at (S2'list)
            do begin
                 Xi=putNextElement(S2'list, Xj)
                 if (Xi <> Null)
                 do begin
                      If (SUi,Y > β)
                      remove Xi from S2'list
                      Xi=putNextElement(S2'list, Xi)
                  End until (Xi==Null)
                  Xj=putNextElement(S2'list, Xj)
            End until (Xj==Null)
            Count πXij,Y
            β'= Update(β)
            repeat (8) – (19) until  optimal πXij,Y
            Sbest=S2'list with optimal πXij,Y value
```

The input node is the actual value set at three locations of the optimal predictor feature (NoP) that results from the PC, SU, and PCSUF selection feature process. We consider the impact on the time lag factor in spatiotemporal based hourly internet-traffic forecasting using RNN. As described in Figure 4, the training and testing phase for t-period forecasting at three locations was performed simultaneously. The RNN model using a hidden layer with a drop-out rate = 0.2 and number of epoch= 1,000.

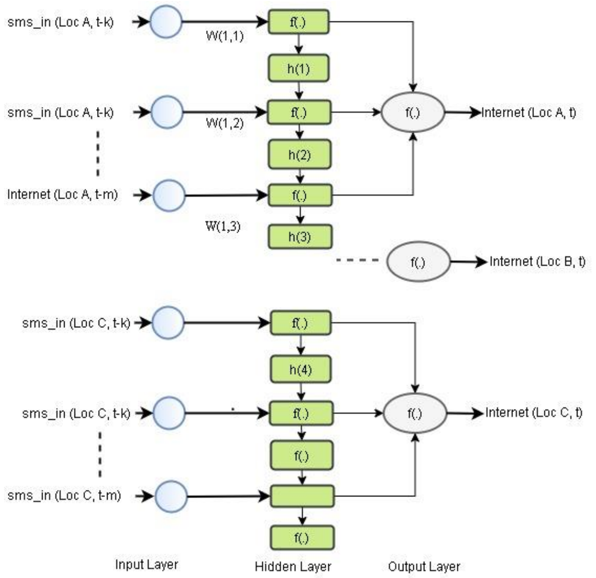

Figure 4.The architecture of RNN internet-traffic forecasting

2.6 Measure of Performance

Through iterations of α and β values, the optimal set of predictor features is obtained. α is the maximum limit of linear feature predictors called redundant features on the PC, and β is the ultimate limit on the SU. Also, such function candidates will be selected using the PC, SU, and PCSUF method in order to receive the appropriate features of predictor built based on the studies [29] and [30]. Formula 8, an extension of the merit value function [30], uses the evaluation function (π) of the optimal predictor subset.

$$\pi_{X_{ij},Y} = \frac{k(\overline{SU})_{Y,X_i}}{k+k(k-1)(\overline{SU})_{X_iX_j}} \quad (8)$$

$\pi_{X_{ij},Y}$ is a merit-value between predictor feature and response variable, $(\overline{SU})_{Y,X_i}$ is the mean SU between inter-correlations, and $(\overline{SU})_{X_iX_j}$ is the mean SU between intra-correlations, and k is the number of predictor features.

The smallest Root Mean Square Error (RMSE), Mean Average Percentage Error (MAPE) and the biggest correlation are the best models (R2). According to Formula 9 and Formula 10, RMSE and MAPE is determined by the difference between the actual data of response variable y_t and the forecasted value $\hat{y}_t$. According to Formula 11, the R^2 is determined by power of the association between the actual data of y_t and the forecasted value of $\hat{y}_t$.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(Y_t-\hat{Y}_t)^2}{n}} \quad (9)$$

$$MAPE = \frac{\sum_{t=1}^{n}\left|\frac{Y_t-\hat{Y}_t}{Y_t}\right|}{n} \quad (10)$$

$$R^2_{y_t,\hat{y}_t} = (cov(y_t,\hat{y}_t)/(std.dev(y_t).stdev(\hat{y}_t))^2 \quad (11)$$

3. Results and Discussions

This section discussed the internet-traffic forecasting model using RNN with one response variable (Y) at three locations for 720 records each location. Furthermore, spatial-temporal RNN modeling was carried out with three predictor variables from three locations (*X1-X3*) through the DPCCA feature extraction process, selection, and removal redundancy using PC, SU, and PCSUF. This study tried 5 combinations of methods shown and described in Table 6 to get the best performance.

The flowchart diagram of the overall method used in this study is explained in Figure 5. Spatio-temporal RNN modeling is carried out using DPCCA to extract input features. Using various methods to select and eliminate redundant features to get the optimal predictor feature set (NoP), training and testing are carried out with a proportion of 90%:10% until 50%:50%.

Table 6. Description of Feature Selection and RNN Model

Model	Methods	Description
1	DPCCA-PC-RNN	Spatio-temporal with DPCCA extraction, PC Filter selects linear relevance and removes redundant features and RNN forecasting.
2	DPCCA –SU-RNN	Spatio-temporal with DPCCA extraction, SU Filter selects non-linear relevant and removes redundant features and RNN forecasting.
3	DPCCA-PCSUF-RNN	Spatio-temporal with DPCCA extraction, PCSUF selects relevantly and removes redundant features and RNN forecasting.
4	DPCCA-RNN	Spatio-temporal with DPCCA extraction, without feature selection and RNN forecasting.
5	DPCCA-PCSUF without Remove -RNN	Spatio-temporal with DPCCA extraction, PCSUF selects linear and non-linear relevant features without removing redundant feautures and RNN forecasting.

Testing the filter methods effectiveness in choosing features of predictor to enhance the performance of the forecasting model was carried out using several input variations and the performance's results of the DPCCA extraction, features filter, and RNN implementation for models 1-5 can be seen in Table 7. Five combination models were used to assess whether feature extraction, relevant feature selection, and redundant feature removal and the Spatiotemporal-based RNN would improve the forecasting model's performance compared to one another. This research shows that the second model at Table 7 has the highest $\pi_{X_{ij},Y}$ value and 14 NoP for the internet traffic (Y) forecasting at the three grids. Each grid needs 14 features predictors than all features resulting from DPCCA, which is 42 features (27.6%).

The performance's results of the DPCCA extraction, features filter, and RNN implementation for models 1-5 can be seen in Table 8. The best performing model for

DOI: https://doi.org/10.29207/resti.v7i6.5296

training and testing is to have high values to produce R-square, high correlation, small MAPE, and small RMSE. The best model also has a high merit-value.

The second model also has the best accuracy (R^2), MAPE, and RMSE for RNN forecasting shown in Table 8. This performance and accuracy of forecasting is better than related studies [8].

These results indicate that processing the input from the RNN using the DPCCA-SU (model 2) filter is quite useful. This happens because DPPCA-SU select the features that are needed so that forecasting gets the best performance compared to other combinations of methods. This also shows that the wireless network traffic prediction model is more suitable for using the DPPCA-SU for feature extraction and selection. This is possible because the three predictor features have a spatio-temporal non-linear relationship with internet traffic data.

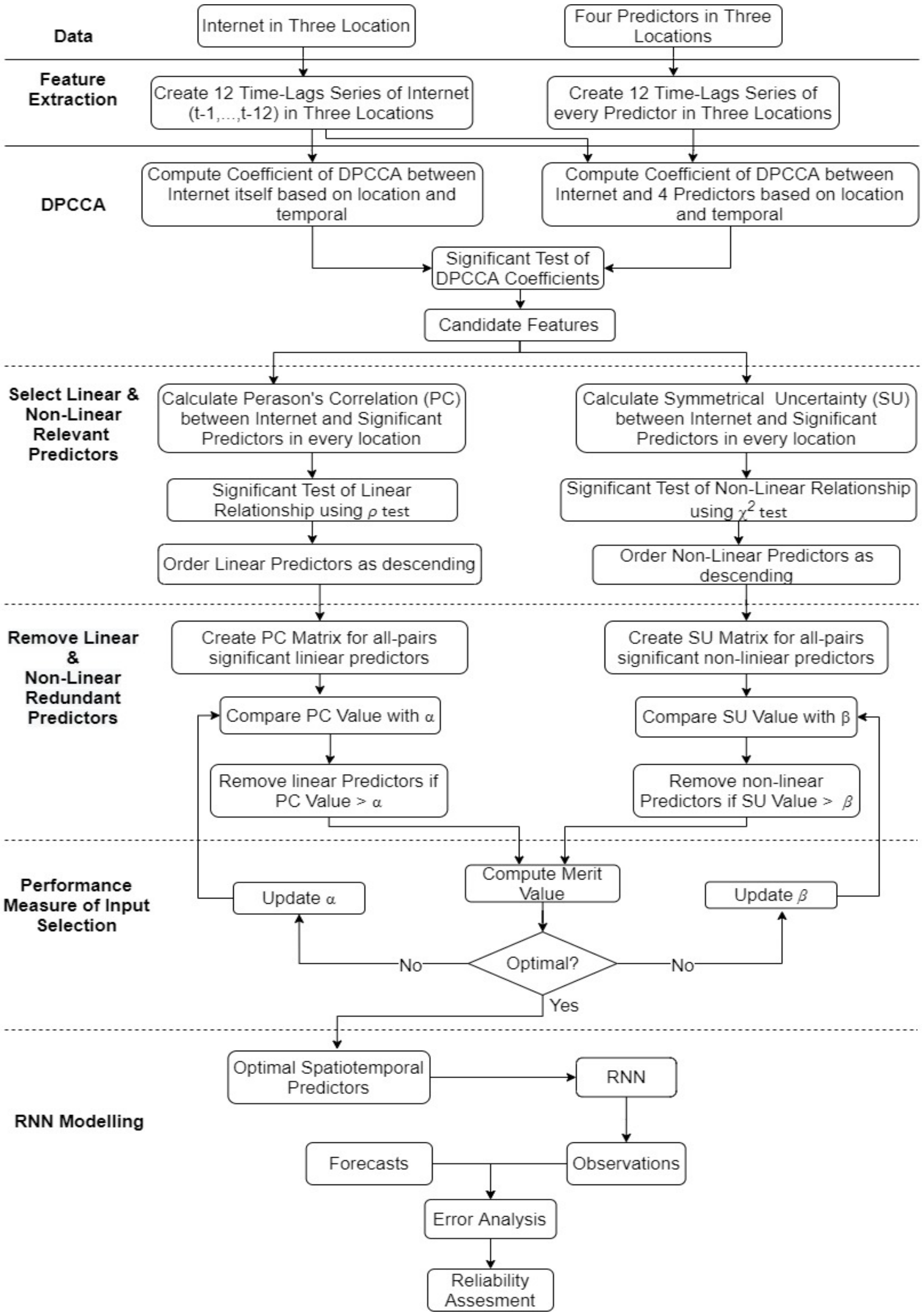

Figure 5.Complete flowchart proposed method

Table 7.Measure of five model performance

Model	NoP	$\pi_{X_{ij},Y}$	Train Test	Train AIC	Train RMSE	Train MAPE (%)	Train R^2	Test RMSE	Test MAPE (%)	Test R^2
1	18	1.42	90:10	-9759.25	938.8	20	0.65	1001.9	25	0.60
			80:20	-8821.12	479.4	12	0.93	607.2	16	0.92
			70:30	-9027.24	630.0	14	0.90	831.3	20	0.86
			60:40	-8262.55	326.4	10	0.96	485.3	13	0.92
			50:50	-8203.56	312.6	10	0.96	493.8	13	0.91
2	14	**1.43**	90:10	-8055.77	339.9	9	0.96	518.1	14	0.95
			80:20	-7406.15	177.1	6	0.99	372.3	10	0.97
			70:30	-7547.92	194.5	7	0.99	435.2	12	0.97
			60:40	-7418.35	178.3	6	0.99	335.1	10	0.98
			50:50	-7434.77	180.8	6	0.99	**243.0**	**7**	**0.98**
3	27	0.34	90:10	-8027.58	278.7	8	0.97	454.6	13	0.93
			80:20	-9125.84	674.6	14	0.82	850.5	20	0.82
			70:30	-8532.48	417.2	11	0.95	602.7	15	0.91
			60:40	-7998.19	272.8	8	0.97	418.7	11	0.94
			50:50	-7994.26	271.2	8	0.97	411.0	10	0.93
4	169	0.22	90:10	-7495.97	254.2	7	0.98	486.5	14	0.92
			80:20	-8098.32	403.2	10	0.95	607.9	16	0.92
			70:30	-8121.91	411.7	10	0.95	681.1	18	0.90
			60:40	-8016.77	380.8	10	0.95	549.7	14	0.90
			50:50	-7365.56	234.7	7	0.98	430.2	11	0.92
5	172	0.27	90:10	-8138.46	420.6	10	0.95	609.4	15	0.90
			80:20	-8121.26	410.1	10	0.96	654.2	17	0.92
			70:30	-7432.38	246.1	7	0.98	509.6	13	0.92
			60:40	-7413.82	241.1	8	0.98	434.0	11	0.93
			50:50	-7423.87	241.5	7	0.98	440.9	11	0.92

Table 8.Model performance of RNN

Model	NoP	$\pi_{X_{ij},Y}$	Train AIC	Train RMSE	Train MAPE (%)	Train R^2	Test RMSE	Test MAPE (%)	Test R^2
1	18	1.432	-8814.75	537.4	13	0.88	683.9	17	0.84
2	14	1.425	-7572.59	214.1	7	0.98	**380.7**	**10**	**0.97**
3	27	0.335	-8335.67	382.9	10	0.94	547.5	14	0.91
4	169	0.217	-7819.71	336.9	9	0.96	551.1	15	0.91
5	172	0.272	-7705.96	311.9	9	0.97	529.6	14	0.92

4. Conclusion

Based on the outcomes of implementation and experimental results for five variations of the model, several conclusions can be drawn. Firstly, the best performance for internet-traffic forecasting with the RNN is achieved by employing feature extraction using the DPCCA method and non-linear feature selection (SU). Secondly, it is evident that the number of predictor inputs alone does not guarantee improved model performance, as demonstrated by Models 4 and 5, which lacked feature selection or redundant removal. Lastly, the hybrid model incorporating RNN and feature selection showcases high performance due to the inclusion of a spatial factor and feature selection mechanism. These findings highlight the importance of utilizing advanced techniques such as DPCCA, non-linear feature selection, and hybrid models to enhance internet-traffic forecasting with RNN.

References

[1] Y. Liu, B. Li, X. Sun, and Z. Zhou, "A fusion model of SWT, QGA and BP neural network for wireless network traffic prediction," *Int. Conf. Commun. Technol. Proceedings, ICCT*, pp. 769–774, 2013, doi: 10.1109/ICCT.2013.6820478.

[2] L. G, W. Q, and L. H, "Network traffic prediction based on wavelet neural network improved by genetic algorithm," *Int. J. Inf.*, vol. 16, no. 2, pp. 1047–1052, 2013.

[3] J. H. Faghmous and V. Kumar, "Spatio-temporal Data Mining for Climate Data: Advances, Challenges, and Opportunities," *Stud. Big Data*, vol. 1, pp. 83–116, 2014, doi: 10.1007/978-3-642-40837-3_3.

[4] S. Mukherjee, R. Ray, R. Samanta, M. H. Khondekar, and G. Sanyal, "Nonlinearity and chaos in wireless network traffic," *Chaos, Solitons & Fractals*, vol. 96, pp. 23–29, Mar. 2017, doi: 10.1016/J.CHAOS.2017.01.005.

[5] Q. Jiang, X. Yan, and B. Huang, "Performance-Driven Distributed PCA Process Monitoring Based on Fault-Relevant Variable Selection and Bayesian Inference," *IEEE Trans. Ind. Electron.*, vol. 63, no. 1, pp. 377–386, Jan. 2016, doi: 10.1109/TIE.2015.2466557.

[6] L. Nie, D. Jiang, S. Yu, and H. Song, "Network traffic prediction based on deep belief network in wireless mesh backbone networks," *IEEE Wirel. Commun. Netw. Conf. WCNC*, May 2017, doi: 10.1109/WCNC.2017.7925498.

[7] C. Zhang, H. Zhang, D. Yuan, and M. Zhang, "Citywide Cellular Traffic Prediction Based on Densely Connected Convolutional Neural Networks," *IEEE Commun. Lett.*, vol. 22, no. 8, pp. 1656–1659, Aug. 2018, doi: 10.1109/LCOMM.2018.2841832.

[8] M. Li, Y. Wang, Z. Wang, and H. Zheng, "A deep learning method based on an attention mechanism for wireless network traffic prediction," *Ad Hoc Networks*, vol. 107, p. 102258, Oct. 2020, doi: 10.1016/J.ADHOC.2020.102258.

[9] G. L. Santos, P. Rosati, T. Lynn, J. Kelner, D. Sadok, and P. T. Endo, "Predicting Short-term Mobile Internet Traffic from

Internet Activity using Recurrent Neural Networks," *Int. J. Netw. Manag.*, vol. 32, no. 3, Oct. 2020, doi: 10.1002/nem.2191.

[10] G. Siracusano and A. La Corte, "Forecasting Mobile Traffic with Spatiotemporal correlation using Deep Regression," Jul. 2019, Accessed: Jul. 05, 2023. [Online]. Available: https://arxiv.org/abs/1907.10865v1

[11] D. Zhang, L. Liu, C. Xie, B. Yang, and Q. Liu, "Citywide Cellular Traffic Prediction Based on a Hybrid Spatiotemporal Network," *Algorithms 2020, Vol. 13, Page 20*, vol. 13, no. 1, p. 20, Jan. 2020, doi: 10.3390/A13010020.

[12] L. Wang, Z. Wang, and S. Liu, "An effective multivariate time series classification approach using echo state network and adaptive differential evolution algorithm," *Expert Syst. Appl.*, vol. 43, pp. 237–249, Jan. 2016, doi: 10.1016/J.ESWA.2015.08.055.

[13] A. Bagnall, J. Lines, A. Bostrom, J. Large, and E. Keogh, "The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances," *Data Min. Knowl. Discov.*, vol. 31, no. 3, pp. 606–660, May 2017, doi: 10.1007/S10618-016-0483-9/FIGURES/17.

[14] G. Barlacchi *et al.*, "A multi-source dataset of urban life in the city of Milan and the Province of Trentino," *Sci. Data 2015 21*, vol. 2, no. 1, pp. 1–15, Oct. 2015, doi: 10.1038/sdata.2015.55.

[15] S. Astutik, N. Iriawan, Suhartono, and Sutikno, "Hybrid state-space model and adjusting procedure based on Bayesian approaches for spatio-temporal rainfall disaggregation," *ICSSBE 2012 - Proceedings, 2012 Int. Conf. Stat. Sci. Bus. Eng. "Empowering Decis. Mak. with Stat. Sci.*, pp. 24–27, 2012, doi: 10.1109/ICSSBE.2012.6396520.

[16] M. E. Banihabib, A. Ahmadian, and F. S. Jamali, "Hybrid DARIMA-NARX model for forecasting long-term daily inflow to Dez reservoir using the North Atlantic Oscillation (NAO) and rainfall data," *GeoResJ*, vol. 13, pp. 9–16, Jun. 2017, doi: 10.1016/J.GRJ.2016.12.002.

[17] B. Bonev, F. Escolano, and M. Cazorla, "Feature selection, mutual information, and the classification of high-dimensional patterns: Applications to image classification and microarray data analysis," *Pattern Anal. Appl.*, vol. 11, no. 3–4, pp. 309–319, Sep. 2008, doi: 10.1007/S10044-008-0107-0/METRICS.

[18] E. Y. Arini, R. Hidayat, and A. Faqih, "Rainfall Simulation Using RegCM4 Model in Kalimantan during El Nino Southern Oscillation," *Procedia Environ. Sci.*, vol. 24, pp. 70–86, Jan. 2015, doi: 10.1016/J.PROENV.2015.03.011.

[19] A. Saikhu, A. Z. Arifin, and C. Fatichah, "Non-Linear Spatio-Temporal Input Selection for Rainfall Forecasting Using Recurrent Neural Networks," *Proceeding - 2018 Int. Semin. Intell. Technol. Its Appl. ISITIA 2018*, pp. 351–356, Jul. 2018, doi: 10.1109/ISITIA.2018.8710864.

[20] X.-Y. Qian, Y.-M. Liu, Z.-Q. Jiang, B. Podobnik, W.-X. Zhou, and H. E. Stanley, "Detrended partial cross-correlation analysis of two nonstationary time series influenced by common external forces," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 91, no. 6, Apr. 2015, doi: 10.1103/PhysRevE.91.062816.

[21] N. Yuan, Z. Fu, H. Zhang, L. Piao, E. Xoplaki, and J. Luterbacher, "Detrended Partial-Cross-Correlation Analysis: A New Method for Analyzing Correlations in Complex System," *Sci. Reports 2015 51*, vol. 5, no. 1, pp. 1–7, Jan. 2015, doi: 10.1038/srep08143.

[22] C. Shen, "Analysis of detrended time-lagged cross-correlation between two nonstationary time series," *Phys. Lett. A*, vol. 379, no. 7, pp. 680–687, Mar. 2015, doi: 10.1016/J.PHYSLETA.2014.12.036.

[23] J. Tang, S. Alelyani, and H. Liu, "Feature selection for classification: A review," *Data Classif. Algorithms Appl.*, pp. 37–64, Jan. 2014, doi: 10.1201/B17320.

[24] M. K. Ebrahimpour and M. Eftekhari, "Ensemble of feature selection methods: A hesitant fuzzy sets approach," *Appl. Soft Comput.*, vol. 50, pp. 300–312, Jan. 2017, doi: 10.1016/J.ASOC.2016.11.021.

[25] Y. Yu, M. Wang, L. Lin, and C. Zhang, "Human Motion Sequence Recognition Based on Feature Selection and Support Vector Machine," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 646, no. 1, p. 012012, Oct. 2019, doi: 10.1088/1757-899X/646/1/012012.

[26] K. Mohammadi, S. Shamshirband, S. Motamedi, D. Petković, R. Hashim, and M. Gocic, "Extreme learning machine based prediction of daily dew point temperature," *Comput. Electron. Agric.*, vol. 117, pp. 214–225, Sep. 2015, doi: 10.1016/J.COMPAG.2015.08.008.

[27] N. Ngoc Thach *et al.*, "Spatial pattern assessment of tropical forest fire danger at Thuan Chau area (Vietnam) using GIS-based advanced machine learning algorithms: A comparative study," *Ecol. Inform.*, vol. 46, pp. 74–85, Jul. 2018, doi: 10.1016/J.ECOINF.2018.05.009.

[28] N. D. C. Lewis, *Neural networks for time series forecasting with R : an intuitive step by step blueprint for beginners*.

[29] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training Recurrent Neural Networks," *30th Int. Conf. Mach. Learn. ICML 2013*, no. PART 3, pp. 2347–2355, Nov. 2012, Accessed: Jul. 05, 2023. [Online]. Available: https://arxiv.org/abs/1211.5063v2

[30] N. Challita, M. Khalil, and P. Beauseroy, "New technique for feature selection: Combination between elastic net and relief," *2015 3rd Int. Conf. Technol. Adv. Electr. Electron. Comput. Eng. TAEECE 2015*, pp. 262–267, May 2015, doi: 10.1109/TAEECE.2015.7113637.

DOI: https://doi.org/10.29207/resti.v7i6.5296