

Article

Adaptive Modeling of Prediction of Telecommunications Network Throughput Performances in the Domain of Motorway Coverage

Milorad K. Banjanin ¹, Mirko Stojčić ^{2,*}, Dejan Drajić ³, Zoran Ćurguz ², Zoran Milanović ⁴ and Aleksandar Stjepanović ²

- ¹ Department of Computer Science and Systems, Faculty of Philosophy Pale, University of East Sarajevo, Alekse Šantića 1, 71420 Pale, Bosnia and Herzegovina; milorad.banjanin@ff.ues.rs.ba
- ² Department of Information and Communication Systems in Traffic, Faculty of Transport and Traffic Engineering Dobo, University of East Sarajevo, Vojvode Mišića 52, 74000 Dobo, Bosnia and Herzegovina; zoran.curguz@sf.ues.rs.ba (Z.Ć.); aleksandar.stjepanovic@sf.ues.rs.ba (A.S.)
- ³ Department of Telecommunications, Faculty of Electrical Engineering, University of Belgrade, Bulevar Kralja Aleksandra 73, 11120 Beograd, Serbia; ddrajic@etf.bg.ac.rs
- ⁴ Forensics Department, University of Criminal Investigation and Police Studies, Cara Dušana 196, 11080 Zemun, Serbia; zoran.milanovic@kpu.edu.rs
- * Correspondence: mirko.stojcic@sf.ues.rs.ba

Abstract: The main goal of this paper is to create an adaptive model based on multilayer perceptron (MLP) for prediction of average downlink (DL) data throughput per user and average DL data throughput per cell within an LTE network technology and in a geo-space that includes a segment of the Motorway 9th January with the access roads. The accuracy of model prediction is estimated based on relative error (RE). With multiple trainings and testing of 30 different variants of the MLP model, with different metaparameters the final model was chosen whose average accuracy for the Cell Downlink Average Throughput variable is 89.6% (RE = 0.104), while for the Average User Downlink Throughput variable the average accuracy is 88% (RE = 0.120). If the coefficient of determination is observed, the results showed that the accuracy of the best selected prediction model for the first variable is 1.4% higher than the accuracy of the prediction of the selected model for the second dependent variable. In addition, the results showed that the performance of the MLP model expressed over R^2 was significantly better compared to the reference multiple linear regression (MLR) model used.

Keywords: predictive models; MLP; prediction; throughput; LTE; cell; model training; model testing; coefficient of determination



Citation: Banjanin, M.K.; Stojčić, M.; Drajić, D.; Ćurguz, Z.; Milanović, Z.; Stjepanović, A. Adaptive Modeling of Prediction of Telecommunications Network Throughput Performances in the Domain of Motorway Coverage. *Appl. Sci.* **2021**, *11*, 3559. <https://doi.org/10.3390/app11083559>

Academic Editor: Marcos Gestal

Received: 25 March 2021

Accepted: 12 April 2021

Published: 15 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Wireless network technologies in the development of digital communications have reached a strong expansion especially in the past few years. The beginning of this development belonged to mobile devices with limited performance and the ability to connect to the Internet with very low data rates. Today, powerful intelligent telephones and tablets are becoming sensor nodes in a communication network and a kind of generator of user satisfaction not just in the interaction field human-to-human (H2H) but in different communication contexts. The expansion of the interaction field of communication with machine-to-machine (M2M) intelligent components is dynamizing the development of mobile (cellular) networks in the direction of signal coverage and Internet access for mobile users in as large an area as possible, especially in rural areas and along roads. The basic requirements set for operators by users of intelligent devices are their mobility and permanently available Internet connection. In addition, the enormous growth in the amount of different types of data (Big Data), and especially multimedia transmitted over the network,

has increased the awareness of end users about the importance of quality of service (QoS) and quality of user experience (QoE). For this purpose, three key wireless network performance indicators (*Key Performance Indicators*—KPIs) are used: *throughput*, *latency* and *packet loss*. In this paper, the focus is on the research of the network throughput performance, and important throughput instances are *Downlink* (DL) in the direction of *DL Throughput* (DLT), i.e., the average throughput in DL direction *Average DL Throughput* (ADT) used in systems for monitoring, reporting and comparing the quality of network traffic.

The ADT value can be determined physically—continuously or discretely (by time), logically (by service) or geographically (by spatial unit). Most of the generated traffic in the mobile network is the DL traffic, directed from the base station to the user equipment. Thus, in this paper, the variables considered as output (dependent) ones are average DL throughput per individual *cell* of the mobile network and average DL throughput per individual *user* in the network along the selected road in the geo-area of the research. The standard used is *Long Term Evolution* (LTE) which belongs to the fourth generation (4G) of the mobile cellular network technologies. The two observed dependent variables are important indicators of the current state of traffic in the mobile network, whose understanding is important for design and modeling, resource planning and control of operational readiness, functional suitability and reliability of the network, and their values are significantly reflected on the level of the network quality of services (QoS) and on the level of quality of user experience (QoE). Accordingly, the subject of research in this paper is the predictive modeling of variations of the value of the average DL throughput in a mobile network at the cell level and at the user level that will allow us to further define/monitor QoS and QoE.

A segment of the *Motorway 9 January* (M9J), a key road in Republika Srpska, located in the northern part of Bosnia and Herzegovina, was selected for the case study in the geo-space of research. The length of that road segment is 72 km, between the two locations of toll collection (Pay toll)—Jakupovci near Banja Luka (BL) and Johovac near Doboje (Db). In the topological structure with this segment of M9J, the research also includes a part of the road (about 6 km) on the route Jakupovci—BL and also a part of the main road M17, Johovac—Db (about 10 km). The geo-space of the research is covered by the 4G telecommunication network managed by the provider M:tel BL, so the main research goal in this paper is to create an adaptive model based on multilayer perceptron (MLP) for prediction of average downlink (DL) data throughput per user (*Average User Downlink Throughput* in Mbps) and average DL data throughput per cell (*Cell Downlink Average Throughput* in kbps) within the LTE network technology. The initial step in the realized research process was the establishment of institutional cooperation between the scientific research organization—Faculty of Transport and Traffic Engineering Doboje—and the mobile operator company—M:tel Banja Luka—on the basis of the request for availability of the data on classes of realized telecommunication traffic in M:tel network in a certain observation period.

Network performance modeling is a current challenge and problem area that is solved using multiple approaches. In this paper, the MLP machine learning model was chosen because it is characterized by very good predictive abilities due to the possibility of adaptive approximation of nonlinear dependences of the output on the input. The contribution of research is reflected in the fact that providers of telecommunications services need to address more complex issues of their quality, especially in the dimension of end-user satisfaction.

The input data in the created predictive MLP model are presented with 17 independent M:tel network variables in accordance with the theoretical analysis of the research results from the relevant literature and the initial 3GPP standard, whose detailed analysis is given in Section 4 of this paper. Multiple training and testing of 30 different variants of the MLP model, with different combinations of metaparameters, based on relative indicators of prediction accuracy influences the choice of the final model. The quality of the results of this research is checked by comparing the accuracy of the prediction given by the selected MLP model with the performance of the multiple linear regression model (MLR) as one of the

representative models of statistical analysis of the research. In addition to the introduction (Section 1) and the above mentioned Section 4, the syntactic structure of the six sections in this paper consists of theoretical analysis of previous research (Section 2), materials and methods (Section 3), and the main focus next to Section 4 is placed on Section 5 on the presentation of the research results with discussion. Concluding remarks are given in the last, Section 6, followed by an overview of the used literature.

2. Overview of Published Relevant Research

Numerous published studies deal with modeling and analysis of the influence of various factors on the throughput and delay, but also on other key performance indicators in the mobile network where different types of data are transmitted [1]. The paper [2] lists some of the most important factors that affect the throughput, namely radio technology, hardware configuration and limitations, negative effects of signal propagation, user mobility, network infrastructure of the mobile operator. The authors developed a model based on machine learning methods for classifying mobile network stability based on user activity, the mentioned infrastructure, the mobile device model, the link stability, the location, and the day period. Based on empirical research in the paper [3], an analysis of network performance such as throughput, delay, packet loss [4] is given, and it is determined that there is a significant influence of several factors on those variables. An interesting methodology for evaluating the performance of 3G networks and mobile applications is presented in [5]. TCP throughput, Round Trip Time (RTT) and retransmission rate are considered as indicators of network performance, and their dependence on factors such as signal strength, topology, locations, the day period. Research in [6] has shown that there is a correlation between signal strength and throughput, but that it depends on the time of day at which performance measurement was performed.

Modeling of the LTE network performance of 4G networks is the subject of the research of [7]. A similar study was conducted and published by the authors in [8] who developed a stochastic model for throughput prediction in an LTE network. The authors in [9] use the *Random Forest* algorithm to create a model that predicts the throughput in a cellular network based on the operator's data on its performance, radio link quality data and contextual data. A significant result from previous research is made by numerous algorithms for network performance prediction, the systematic comparison of which is presented in [10]. The analysis of latency, as an important indicator of network performance for M2M traffic and online game traffic in the High Speed Packet Access (HSPA) network technology is given in [11], and a similar study was performed in [12].

The model for predicting the level of user experience in web search, based on logistic regression, was created and presented in [13]. For the prediction of the DL and uplink (UL) traffic in the base stations of the mobile network along the highway, a model based on convolutional artificial neural networks is presented in [14]. Using multilayer artificial neural networks in [15], Ur Rehman and his team modeled the DL throughput in an LTE network based on several independent variables related to the radio traffic conditions in the networks. Following the example of that research, with conceptual similarities, in this paper, an MLP model is created in which the metaparameters are completely different, a significantly larger set of independent variables as inputs to the model, but also a significantly larger set of training and testing data. After testing multiple model variants with different training–testing ratios, it was found that a training and testing data ratio of 70%:30% gives good results, although a 50%:50% ratio also gives good model results. From the practical aspect, the purposeful application of the MLP model in a specific geo-space and analytically explained interactions of virtual and physical devices of the M:tel network and the M9J set an additional dimension of the quality of predictive traffic modeling in networks. The following Table 1 presents an overview of the referenced previous research in the subject of this paper with highlighted novelties presented in this paper.

Table 1. Overview of referenced previous research in the subject of this paper.

Number of Reference	Title of the Paper/Book	Applied Model	Keywords That Correlate with This Paper	Novelties in This Paper/Research
[1]	Komunikacioni inženjering	Models of communication theory	Communication, network, performances, model	
[2]	Analyzing Throughput and Stability in Cellular Networks	Bagging based classification, random forest, conditional inference tree	Machine learning model, throughput, prediction, network performance	
[3]	An Empirical Characterization of Cellular Network Performance	Empirical model	Throughput, 4G LTE, network parameters	
[4]	On the Quest for New KPIs in Mobile Networks: The Impact of Throughput Fluctuations on QoE	Mathematical model	Throughput, KPI, average downlink throughput	
[5]	Anatomizing Application Performance Differences on Smartphones	Empirical model, probability model	Cellular data network performance, throughput	
[6]	Mobile Network Measurements—It's not all about Signal Strength	Statistical model	Throughput, signal strength, network coverage	
[7]	Modeling and Performance Evaluation of LTE Networks with Different TCP Variants	LTE simulation model	LTE network, throughput, TCP performance	-An MLP model with a new set of metaparameters was applied
[8]	A Model for Throughput Prediction for Mobile Users	Stochastic model	LTE user throughput, prediction, mobile networks	-A significantly larger set of independent variables as inputs to the model
[9]	Instantaneous Throughput Prediction in Cellular Networks: Which Information Is Needed?	Random forest	Throughput prediction, cellular networks, users	-A significantly larger set of training and testing data
[10]	An Empirical Study of Throughput Prediction in Mobile Data Networks	Empirical model	TCP throughput, prediction error, mobile networks	-Multiple trainings and testing of 30 different variants of the MLP model, with different metaparameters
[11]	Latency analysis for M2M and Online Gaming traffic in an HSPA network	Statistical model	Latency, traffic, throughput	
[12]	Evaluation of the UTRAN (HSPA) performance in different configurations in the presence of M2M and online gaming traffic	Statistical model	Latency, traffic, throughput	
[13]	Logistic regression based in-service assessment of mobile web browsing service quality acceptability	Binary logistic regression model	QoE prediction, network performance, QoS	
[14]	Mobile Traffic Prediction Based on Densely Connected CNN for Cellular Networks in Highway Scenarios	Convolutional neural networks	Traffic prediction, highway, deep learning	
[15]	LTE Downlink throughput modeling using Neural Networks	Artificial neural network model, deep neural network	Artificial neural networks, user downlink throughput, cell downlink throughput	

3. Materials and Methods

Traffic prediction and planning in existing telecommunications networks is increasingly oriented towards connecting physical and virtual devices based on the use of current concepts of the M2M communication, Wireless Sensor Networks (WSN), Internet of Things (IoT) and Cyber Physical Systems (CPS). Machines or systems, i.e., intelligent devices in the network, communicate with wireless technologies according to the LTE standards, in a way that the devices forward the collected data in the form of packets through the telecommunication network, thus generating traffic, and applications process them and provide information for users. Therefore, traffic generation is performed in the domain of the provider's network, which provides communication service to users—devices, data transfer in the domain of an external or global communication network, and their processing in the application domain. In addition to the growing number of intelligent devices, the diversity of their application is increasingly present in the networks, which affects the setting of various requirements and restrictions related to data transfer speeds, packet sizes, delays, reliability, unattended operation, etc. These communication requirements have been standardized by the Third Generation Partnership Project (3GPP) regulatory body. Therefore, e.g., Technical Specification [16] specifies and establishes the characteristics of the physical layer procedures in the Frequency Division Duplex (FDD) and Time Division Duplex (TDD) modes of Evolved Universal Terrestrial Radio Access (E-UTRA). In [17], the description and definition of the measurements performed by Evolved UMTS Terrestrial Radio Access Network (E-UTRAN) is given. Traffic steering control information is subject to technical specification [18]. Radio Link Control (RLC) protocol specifications are given in [19]. Similar to [17], in [20] the measurement specification performed by E-UTRAN is given. Technical specification [21] is related to improving LTE coverage. Traffic modeling in M2M communications is presented in [22] and [23].

On the other hand, the European Telecommunications Standards Institute (ETSI), which also owns the LTE as the world's dominant mobile phone technology, is in charge of standardizing the abstract Middleware service layer.

For the defined subject and objectives of the research, the methodology of the research process was implemented in several successive steps: 1. Data collection; 2. Creating a predictive model—digital data treatment (defining the model metaparameters, model training, model testing, selection of the final model); 3. Results analysis and discussion; 4. Conclusion.

3.1. Data Collection

Data collection was performed in accordance with the request which specifies the types of data for modeling the average DL network throughput along the M9J in the geo-space of research, at the user level and at the level of individual cells. The data relate mainly to the radio channel characteristics, the physical resource occupancy, the number of users, the base stations and cell parameters, the network topology, and the signal parameters. During the month of January 2021, the mobile operator (M:tel BL) through several iterations delivered a set of the research data provided by a third party, a *vendor*. In the obtained database for the LTE network consisting of a total of 71,053 measurements, the values of the variables were registered in a period of 30 days (from 15 December 2020 to 15 January 2021), with 1-h measurement frequency a day, by individual cells that cover the observed geo-space of the M9J road with the signal. Therefore, each of the 71,053 measurements represents a one-dimensional array or input/output vector consisting of 19 values, of which the first 17 represent independent and the last two, dependent variables.

3.2. Creating a Predictive Model—Digital Data Treatment

Creating a predictive model, based on an artificial neural network (ANN), multilayer Perceptron (MLP), was preceded by structuring a set of the research data into input/output vectors in an Excel file. Out of a total of 71,053 collected input data, the software detected and removed those with unusual values (equal to zero) and with missing values of variables.

Thus, according to the *supervised learning* paradigm, 64,301 inputs/output vectors were included in the finally structured set adapted for training and testing of the predictive model. Supervised learning, as a machine learning technique, implies the existence of a data set of the input x and the corresponding output y , for training so that the created model of classification (y can predict class x) or regression (y can have values in continuity) can predict previously “unseen” inputs.

In LTE networks, as components of modern cyber–physical integrations of physical and digital processes, self-configuration of the software and adaptive devices of machines is performed and maintained by applying corrective decisions [24] on the principles of supervised machine learning. The collection of the input data about the physical processes is performed by sensors, while the information processing of a data set or machine learning of a model is performed by the ANN application through a layered architecture of a series of process elements—interconnected artificial neurons. The way in which an artificial neuron processes the signals is defined by a *transfer function* that is a combination of the *activation* and *output* functions. The activation function maps the individual input signals in a neuron into a single composite signal, while the output function maps the value of the activation function into the output value of the neuron. The most common type of the output function is the *sigmoid function*, which limits the value of the activation function to the interval between 0 and 1. Each synapse, the connection between the neurons, is determined by its weight factor, and during the learning process they are adjusted in learning the function. Modification of these adjustable parameters allows the network to learn a certain law of mapping the vectors from the input space X to the output space $Y = f(X)$. Therefore, it can be said that the ANN is a graph created by composing different transfer functions of individual neurons. Thus, the transfer function is a composite function that maps a vector from the N dimensional input space, also called the *parameter space*, to the neuron *output space* [25]. An artificial neuron whose inputs are assigned the weight factors, and whose output function is a simple step function or threshold function, is called a perceptron, and the problem it most often solves is a linear classification, meaning that data in the space must be linearly separable [26]. The main feature and advantage of the ANN is learning, which involves the process of adapting the network parameters [27].

The fact that MLP, in addition to the input and output layer of the ANN, has one or more hidden layers in its architecture that increase the depth of the network gives it the greatest advantage and power in this class of models because the output function of neurons is nonlinear, usually of the sigmoid shape [28]. According to [29], a multilayer perceptron is considered to allow nonlinear mapping between the input and the corresponding output vector. The learning rate, the number of hidden layers, the number of neurons in each hidden layer, and the shape of the output function of the hidden layer neurons are metaparameters of the model and are estimated before the start of the network training or learning. Therefore, an MLP is designed to approximate (universal approximator) any continuous function, and in addition, it can solve the problems that are not linearly separable. In this paper, the MLP model was created in the IBM SPSS Statistics 22 software that uses nonlinear modeling to detect complex relationships between the collected data. It enables the creation of machine learning models, integration with Big Data and seamless deployment into applications. The process of creating an MLP model based on an initial set of 71,053 input/output vectors, provided by a vendor, can be represented algorithmically in the following successive steps (Figure 1):

- Removal of unusual and missing values to obtain a final set of 64,301 input/output vectors;
- Rescaling the values of 17 input variables;
- Selection of the ratio between the training data set and the testing data set (30%:70%, 40%:60%, 50%:50%, 60%:40%, and 70%:30%);
- Selection of MLP model metaparameters (number of neurons in hidden layers: 10, 10; 20, 20; 30, 30; output function of hidden layer neurons: Hyperbolic tangent or Sigmoid);

- Selection of MLP model training metaparameters (online learning, backpropagation algorithm, initial learning rate = 0.4, lower learning rate limit = 0.001, number of epochs—automatic);
- Selection of training stop criteria (15 min or five iterations without reducing the error function);
- Choice of how to present results (tabular, graphical results);
- Training and testing of MLP model;
- Prediction results: relative error (RE), degree of accuracy, coefficient of determination (R^2);
- For each of the 30 different variants of the MLP model, training and testing are performed three times;
- Selection of the best final model based on the average value for three measurements;
- Selection and presentation of the best individual prediction results for both dependent variables;
- Comparison of MLP model prediction performance with the created MLR model.

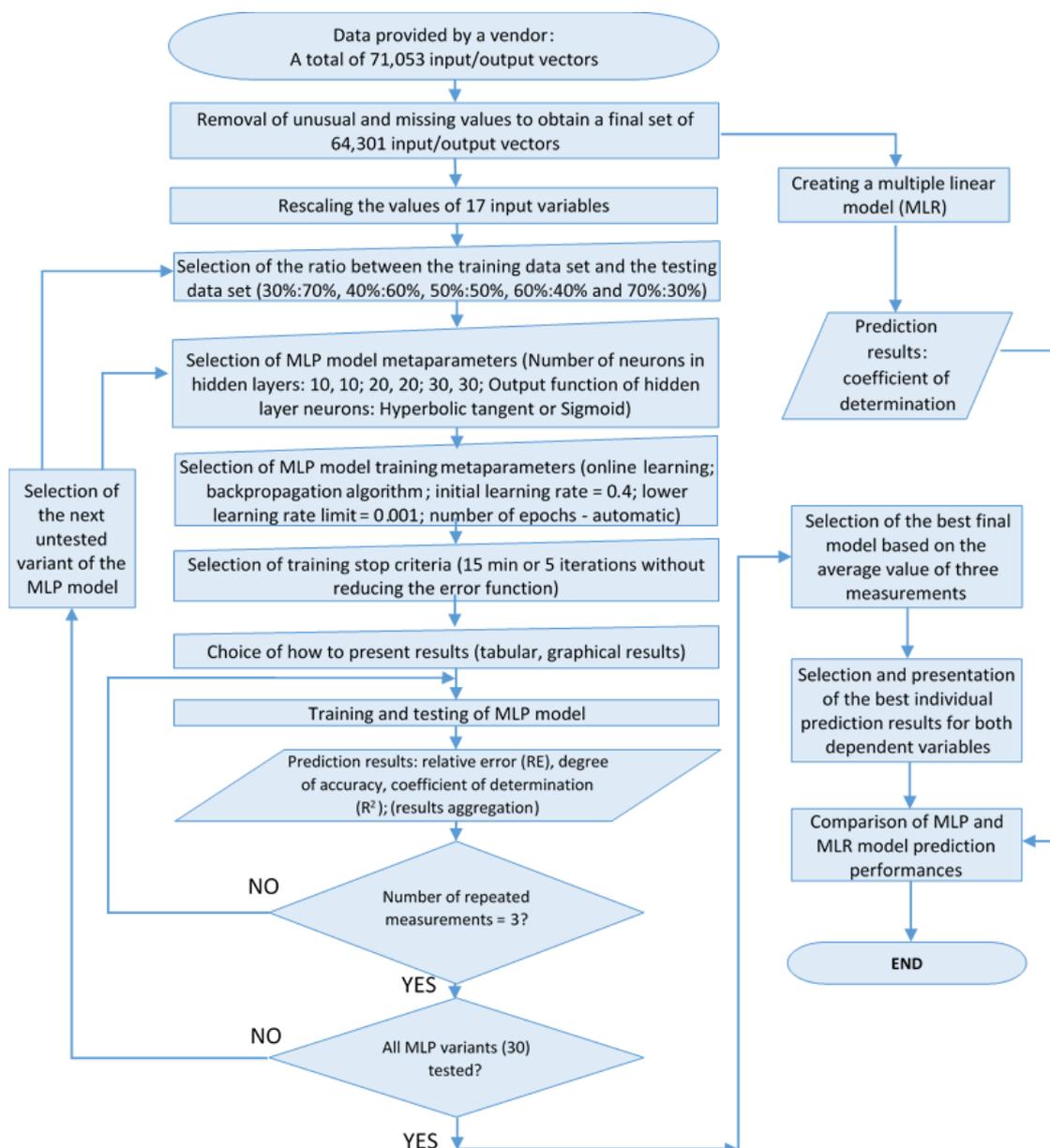


Figure 1. Algorithm for creating MLP model.

According to the above steps of the algorithm, the training and testing of several different MLP models were performed, with different combinations of the metaparameters.

On the data set for testing, the validation of each individual model was performed, i.e., assessment of its quality based on relative criteria: RE and R^2 , which are dimensionless quantities. In general, the relative error can be defined as the ratio of absolute error, which is obtained as the difference between the value of the model prediction and the actual value (real value from the data set), according to the actual value of the observed variable:

$$RE = \left| \frac{\text{Predicted Value} - \text{Real Value}}{\text{Real Value}} \right|. \quad (1)$$

This type of error is relative to the actual value of the variable, has no unit of measure and can be expressed as a percentage. In this study, the two dependent variables, *Cell Downlink Average Throughput* and *Average User Downlink Throughput*, are continuous in nature, meaning that the RE of MLP model training and testing represents the total, average sum of RE values, represented by expression (1), for all measurements. Based on the RE, another parameter of the model quality can be defined, and that is the degree of accuracy of the model which is equal to $1-RE$ (%). Based on the above indicators, the best final model was selected, the results of which are presented in Section 5.

4. Analysis of Research Variables

The permanent increase in the number of connected devices in the network affects the increase in traffic at the cellular level and the entire network, which represents new challenges that service providers must face to ensure a satisfactory level of QoS in the transmission of speech, video and audio streams, images, interactive content [30]. Therefore, it is crucial to perform traffic modeling, which means modeling its parameters, and, in this paper, they represent the two dependent variables: average DL throughput per cell of the mobile network, and average DL throughput per user in the observed network. Along roads in urban and rural areas, the existing transport infrastructure is today rapidly supplemented by modern information and communication technologies (ICT) [31]. Distributed multimedia information systems for traffic monitoring are the subject of research in [32]. Coding algorithms are of great importance in the transmission of multimedia data, and some of them are presented in [33]. Technologies based on the principles of artificial intelligence, such as the adaptive neuro-fuzzy model for traffic signs recognition [34], and for predicting the level of Wi-Fi signals make traffic modeling in such spaces even more important [35].

4.1. Coverage of Deveti Januar Motorway with Mobile LTE Technology

This part of the section provides an overview of the mobile LTE technology coverage in the M:tel network, along the M9J road, from the aspect of the number of the base stations and cells. They are then textually and graphically processed providing a detailed insight into the 17 independent research variables marked as follows: (1) *DL PRB Usage Rate*, (2) *Average CQI*, (3) *RRC.ConnReq.Att*, (4) *RRC.ConnReq.Succ*, (5) *DL ReTrans Rate*, (6) *QPSK.TB.Retrans*, (7) *16QAM.TB.Retrans*, (8) *64QAM.TB.Retrans*, (9) *QPSK.TB*, (10) *16QAM.TB*, (11) *64QAM.TB*, (12) *Traffic.User.Avg*, (13) *DL iBLER*, (14) *DL Cell Traffic Volume*, (15) *QPSK.ErrTB.Ibler*, (16) *16QAM.ErrTB.Ibler*, (17) *64QAM.ErrTB.Ibler*.

In the research process, the collected data, i.e., the values of the independent and dependent variables, as already mentioned, were given with a sampling frequency of 1 h, for a period of 30 days, for each cell that covers M9J with the LTE technology and observed parts of the roads, from the toll booths toward BL and Db. The signal coverage is provided by 28 base stations, and considering that some of them cover up to six cells, and some only one, a network structure of 98 cells was formed along and around the observed road. However, the specific number of the cells that cover the area above the road is smaller, so that in the results of the *Drive Test* measurement of the M:tel operator from 16 December

2020 on the road, 34 cells can be identified which are shown in Figure 2 in different colors (different colors were used to show the coverage and distance between the adjacent cells).

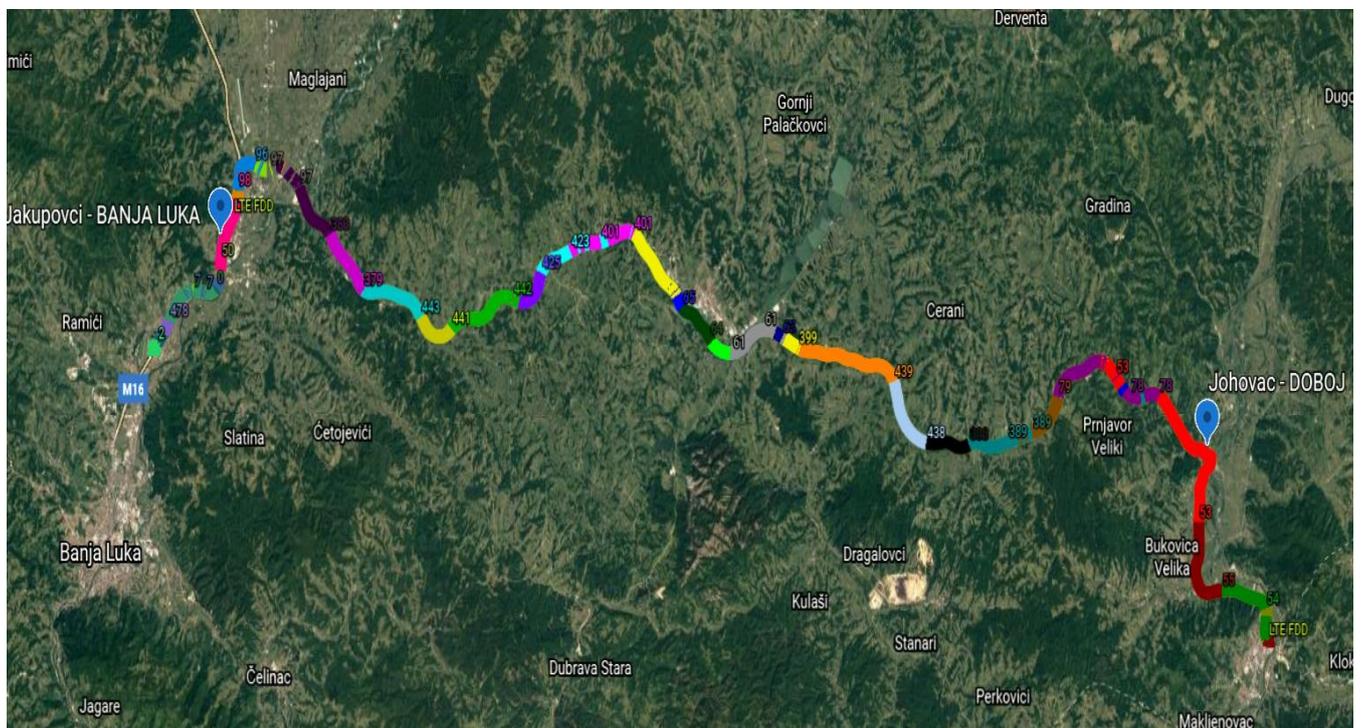


Figure 2. A view at the topology of the coverage of the M9J segment with LTE network technology cells.

In the continuation of this section, eight research variables are analyzed under the points with subheadings of their titles. Integrated under these titles, the other nine independent variables are defined as follows: (1) within Section 4.5, the defined variables are *QPSK.TB.Retrans*, *16QAM.TB.Retrans*, *64QAM.TB.Retrans*, *QPSK.TB*, *16QAM.TB*, and *64QAM.TB* due to the direct connection to the retransmission rate—*DL Retrans Rate* which is also presented mathematically within this subtitle; (2) regarding that the variables *QPSK.ErrTB.Ibler*, *16QAM.ErrTB.Ibler* and *64QAM.ErrTB.Ibler* influence *Initial Block Error Rate (iBLER)*, they are analyzed within Section 4.7.

4.2. DL PRB Usage Rate

The Physical Resource Block (PRB) is the smallest unit of resource that can be allocated to a user in an LTE network. If the frequency domain is observed, the resource blocks consist of 12 subcarriers of 15 kHz each or 24 subcarriers of 7.5 kHz each. This means that the bandwidth of the PRB is 180 KHz. In most cases, the number of subcarriers used per block of resources is 12 [36].

In the time domain, one physical block of resources occupies one slot (seven symbols) in duration of 0.5 ms, which represents 1/20 of the structure of the radio frame used for DL and UL data transmission in the LTE [37]. The smaller resource unit and the smallest discrete part of the radio frame is a resource element (RE), which has the dimensions 1 symbol \times 1 applicant, which means that one PRB has 84 resource elements. The described structure of one physical block of resources with the resource elements is shown in Figure 3.

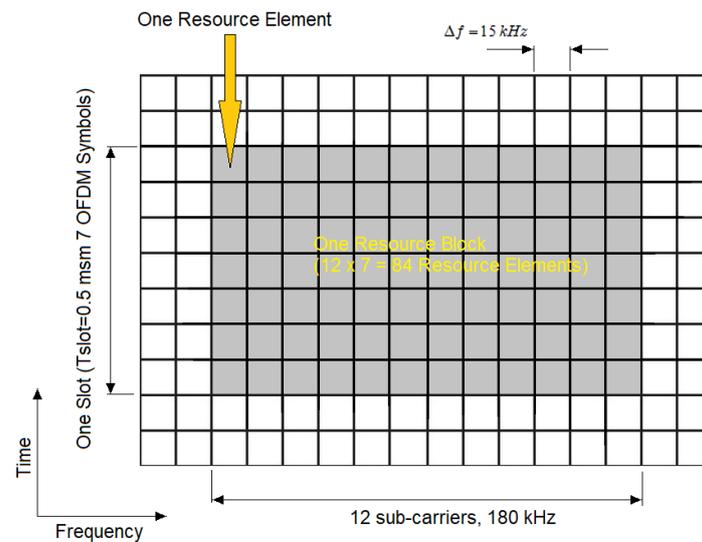


Figure 3. Structure of physical block of resource elements (adapted from [36]).

The independent variable or a counter labeled *DL PRB Usage Rate* represents the average utilization of DL physical blocks of resources (expressed as a percentage). It is calculated as the ratio of the average number of used physical blocks of the physical DL shared channel (*Physical Downlink Shared Channel—PDSCH* [38]) and the total available number of the DL physical resource blocks multiplied by 100. PDSCH is a physical DL common transport channel which, in addition to transmitting the user data, has the function of transmitting the data important for control, and specific information of higher layers and DL system information.

4.3. Average CQI

In an LTE network, the value of the *Channel Quality Indicator (CQI)* is sent by the mobile device via the UL connection to the eNodeB, which provides the *User Equipment (UE)* connection to the rest of the mobile network, and which, based on that value for data transmission in the DL direction, chooses the appropriate *Modulation and Coding Scheme (MCS)* (Table 2).

Table 2. CQI values [39].

CQI Index	Modulation	Approximate Code Rate	Efficiency (Information Bits per Symbol)
0	No transmission	-	-
1	QPSK	0.076	0.1523
2	QPSK	0.12	0.2344
3	QPSK	0.19	0.3770
4	QPSK	0.30	0.6016
5	QPSK	0.44	0.8770
6	QPSK	0.59	1.1758
7	QPSK	0.37	1.4766
8	16QAM	0.48	1.9141
9	16QAM	0.60	2.4063
10	16QAM	0.45	2.7305
11	64QAM	0.55	3.3223
12	64QAM	0.65	3.9023
13	64QAM	0.75	4.5234
14	64QAM	0.85	5.1152
15	64QAM	0.93	5.5547

In this way, it defines the data transfer rate in the channel. The CQI value is determined based on the signal-to-interference plus noise (SINR) ratio and based on the characteristics of the mobile device. Most often, the mobile device selects the CQI value, and thus the MCS, so that the error rate (BLER) does not exceed 10%. Therefore, each CQI value is mapped to the corresponding modulation scheme (Quadrature Phase Shift Keying—QPSK, Quadrature Amplitude Modulation—QAM: 16QAM and 64QAM) which corresponds to a certain coding rate (Code Rate), and the number of bits per symbol. Table 2 shows CQI values according to 3GPP LTE standards [39].

4.4. *RRC.ConnReq.Att* and *RRC.ConnReq.Succ*

Radio resource control (RRC) is a multirole network protocol that defines communication between an eNodeB and a mobile device. From a base station perspective (eNodeB), a mobile device in a cell can be found in two basic states [40]: *Idle Mode* and *Connected Mode*.

The independent variable *RRC.ConnReq.Att* represents the number of attempts by the UE to establish a connection with an eNodeB, i.e., to start the RRC connection setup procedure, as shown in Figure 4 under number 1. Each time the procedure is started, this variable is incremented by one. In this case, the number of repeated connection–retransmission is not counted. *RRC.ConnReq.Succ* represents the number of confirmation messages about the completion of the Connection Setup procedure that the eNodeB receives from the user equipment in the cell and is incremented by one each time the eNodeB receives this message. In Figure 4, this step is marked under number 3.

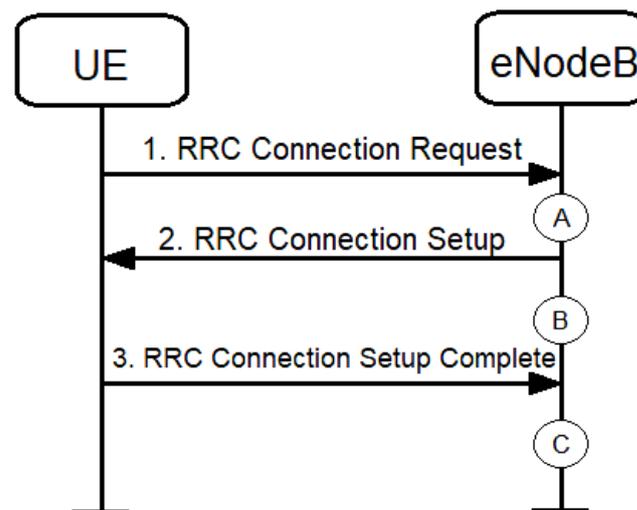


Figure 4. RRC connection setup procedure (adapted from [41]).

4.5. *DL ReTrans Rate*

As mentioned above, DL-SCH is a shared transport channel for transmitting user data in the DL direction. If the connection between the eNodeB and the user equipment is not established in one attempt, a reattempt is undertaken to send the packet over this channel or the retransmission, usually based on the modulation and coding scheme determined during the initial transmission. The data packets are sent in *Transport Blocks (TB)*, and the block size is determined by the number of bits sent within one *Transmission Time Interval (TTI)* lasting 1 ms. If retransmissions by signal modulation types are taken into account, the value of the retransmission rate can be calculated according to the following expression:

$$\text{DL ReTrans Rate} = \frac{\text{QPSK.TB.Retans} + 16\text{QAM.TB.Retans} + 64\text{QAM.TB.Retans}}{\text{QPSK.TB} + 16\text{QAM.TB} + 64\text{QAM.TB} + \text{QPSK.TB.Retans} + 16\text{QAM.TB.Retans} + 64\text{QAM.TB.Retans}} \times 100 \quad (2)$$

where the meanings of the individual variables that appear in the expression (2) are given in Table 3, and they are, in addition to *DL ReTrans Rate*, used as independent variables in the predictive model.

Table 3. Values by which the DL retransmission rate is calculated.

Mark	Meaning
QPSK.TB.Retrans	No of retransmitted TB into DL-SCH under QPSK modulation
16QAM.TB.Retrans	No of retransmitted TB into DL-SCH under 16QAM modulation
64QAM.TB.Retrans	No of retransmitted TB into DL-SCH under 64QAM modulation
QPSK.TB	No of initially emitted TB into DL-SCH under QPSK modulation
16QAM.TB	No of initially emitted TB into DL-SCH under 16QAM modulation
64QAM.TB	No of initially emitted TB into DL-SCH under 64QAM modulation

4.6. Traffic.User.Avg

The value of this variable depends on the user states at the time of sampling (idle state or connected state). Every second, the number of UEs in the connected state in the cell is registered. At the end of the measurement interval, the average number of UEs in the specified state is taken as the value of the independent variable marked with *Traffic.User.Avg*.

4.7. DL iBLER

The term *Block Error Rate* (BLER) can be defined as the ratio of transmitted blocks with error and the total number of transferred blocks and is expressed as a percentage. Detection of errors in the transport blocks on the receiver side is performed using the *Cyclic Redundancy Check* (CRC) technique. In measuring network performance, the *Initial Block Error Rate* (iBLER) metric is often used, which shows the ratio of the number of blocks with initial transmission errors and the total number of initially transmitted transport blocks [42]. If certain types of modulation are observed in the transport channel, then the expression for iBLER can be written in the following form:

$$\text{DL iBLER} = \frac{\text{QPSK.ErrTB.Ibler} + 16\text{QAM.Err.TB.Ibler} + 64\text{QAM.ErrTB.Ibler}}{\text{QPSK.TB} + 16\text{QAM.TB} + 64\text{QAM.TB}} \quad (3)$$

where the meanings of the marks of the variables appearing in the model are defined in Table 4. The three variables shown are, in addition to iBLER, also taken as independent variables of the predictive model.

Table 4. Values by which the iBLER value is calculated.

Mark	Meaning
QPSK.ErrTB.Ibler	Number TB with initial errors under QPSK modulation
16QAM.ErrTB.Ibler	Number TB with initial errors under 16QAM modulation
64QAM.ErrTB.Ibler	Number TB with initial errors under 64QAM modulation

4.8. DL Cell Traffic Volume

This variable represents the total aggregated DL traffic in the cell up to the moment of sampling and is expressed in Gbit. According to the standard 3GPP TS 23.203 rel 8, the total aggregate traffic is grouped into nine classes by the quality of service, so that the *QoS Class Identifier* (QCI) clearly identifies each traffic class with numbers from 1 to 9, which is the case in the M:tel LTE network. QCI definitions and explanations are given in [43,44]. Each QCI defines the following four parameters, which are given in Table 5 in addition to the service examples for each QCI value [45]:

1. Bit rate: guaranteed Bit Rate (GBR) or Non-Guaranteed Bit Rate (Non-GBR);
2. Priority level: a number in the range between 1 and 9, where 1 represents the highest priority;
3. Packet Delay Budget—PDB. Defines the upper limit of the allowed packet delay;
4. Packet Error and Loss Rate (PELR). Defined as the ratio of the number of packets with errors-lost packets and the total number of sent packets [46].

The examples of services are given in the following Table 5 and can be classified into traffic classes in a 3G mobile network, according to the definition of the Organization for the Development of International Telecommunication Standards (3GPP), by QoS criteria: (1) conversational traffic (voice transmission—talks and lectures); (2) streaming traffic (audio/video streaming—video telephony, telnet, voice and video services); (3) interactive traffic (internet search—interactive applications such as interactive email and interactive web browsing); (4) background traffic (W3, email—background download, FTP, news and Telnet, download of background files). This clustering is usually used as a base in network traffic.

The parameters listed in Table 5, among others, define the quality of an individual service from the aspect of the mobile operator, i.e., the service provider. Although the QoS is a well-defined and established research domain, it is often necessary, but also difficult, to distinguish it from the quality of the user experience. The QoE puts the person-user and his perception of the quality of experience in the foreground, but in general it can be said that with the increase of QoS, the operator can expect a higher value of QoE [47].

Table 5. QCI values and corresponding parameters in LTE network [45].

QCI	Priority	Resource Type	Delay Budget (ms)	PELR	Example of Service
1	2	GBR	100	10^{-2}	Conversational voice
2	4	GBR	150	10^{-3}	Conversational video (live streaming)
3	3	GBR	50	10^{-3}	Real-time gaming
4	5	GBR	300	10^{-6}	Nonconversational video (buffered streaming)
5	1	Non-GBR	100	10^{-6}	IMS signaling
6	6	Non-GBR	300	10^{-6}	Video (buffered streaming), TCP-based (e.g., www, e-mail, chat, FTP, P2P file sharing, progressive video, etc.)
7	7	Non-GBR	100	10^{-3}	Voice, video, interactive gaming
8	8	Non-GBR	300	10^{-6}	Video (buffered streaming), TCP-based (e.g., www, e-mail, chat, FTP, P2P file sharing, progressive video, etc.)
9	9	Non-GBR	300	10^{-6}	Video (buffered streaming), TCP-based (e.g., www, e-mail, chat, FTP, P2P file sharing, progressive video, etc.)

5. Results and Discussion

This section presents the research results and the performance of the created multiple linear regression and MLP model in order to determine the justification for selection of a better model for traffic prediction in networks. Both models are created in IBM SPSS technology. It is known from the theoretical analysis of previous research that *linear regression models* belong to the group of statistical models and as such have certain shortcomings when it comes to their application on large databases. At the same time, machine learning-based predictive models, such as MLP, show better agreement (model fit) with available data because an optimal prediction solution is found through an iterative parameter setting procedure [27].

5.1. Model of Multiple Linear Regression (MLR)

The multiple linear regression model, as one of the most commonly used statistical models, was also created in the IBM SPSS technology. Initially, in the regression analysis, the dimensionality of the input space was represented by 17 previously analyzed independent variables. Unlike MLP, the output space in the regression model consists of only one dependent variable, so that predicting the value of the average data throughput at the cell level and at the user level along the observed segment of the motorway requires two independent models.

By regression analysis, the variable *RRC.ConnReq.Att* was excluded from both models due to the existence of strong multicollinearity with the variable *RRC.ConnReq.Succ*, which means that in this case it represents a redundancy. Table 6 shows the results of the regression analysis that allow the value of the observed target dependent variable (Y_1 and Y_2) to be

calculated based on a linear combination of inputs— independent variables denoted by $X_1, X_2, X_3, \dots, X_{16}$. Of particular interest is the column showing the nonstandardized coefficients (B) that appear in the multiple linear model. Coefficients B represent the estimated model parameters from the data using the method of least squares. Based on the values shown in the column marked with Sig., it can be concluded that all coefficients, in both models, including the constant have a value less than 0.05, which means that they are statistically significant and need to be included in the models.

Based on the coefficients shown in Table 6, the following regression model was formed, which represents a linear dependence of the value of the variable *Cell Downlink Average Throughput* of the 16 research variables:

$$Y_1 = -922.443 - 416.156X_1 + 1359.353X_2 + 0.428X_3 - 22.248X_4 - 289.941X_5 + 748.020X_6 + 1043.475X_7 - 0.157X_8 - 0.046X_9 + 0.033X_{10} + 0.075X_{11} + 0.016X_{12} - 0.025X_{13} + 0.004X_{14} - 0.005X_{15} + 0.003X_{16} \tag{4}$$

The value of the determination coefficient of this model is $R^2 = 0.541$.

Table 6. Basic results of the regression analysis.

Dependent Variables Independent Variables	Cell Downlink Average Throughput (Y_1)			Average User Downlink Throughput (Y_2)		
	Unstandardized Coefficients		Sig.	Unstandardized Coefficients		Sig.
	B	Std. Error		B	Std. Error	
(Constant)	-922.443	214.142	0.000	-11.194	0.358	0.000
DL PRB Usage Rate (X_1)	-416.156	4.679	0.000	-0.941	0.008	0.000
Average CQI (X_2)	1359.353	19.397	0.000	3.261	0.032	0.000
RRC.ConnReq.Succ (X_3)	0.428	0.033	0.000	0.001	0.000	0.000
Traffic.User.Avg (X_4)	-22.248	5.621	0.000	-0.103	0.009	0.000
DL ReTrans Rate (X_5)	-289.941	11.376	0.000	0.202	0.019	0.000
DL iBLER (X_6)	748.020	12.815	0.000	0.046	0.021	0.033
DL Cell Traffic Volume (X_7)	1043.475	6.101	0.000	1.031	0.010	0.000
16QAM.ErrTB.Ibler (X_8)	-0.157	0.005	0.000	-7.331×10^{-5}	0.000	0.000
64QAM.ErrTB.Ibler (X_9)	-0.046	0.003	0.000	7.704×10^{-5}	0.000	0.000
QPSK.ErrTB.Ibler (X_{10})	0.033	0.003	0.000	4.820×10^{-5}	0.000	0.000
16QAM.TB.Retrans (X_{11})	0.075	0.004	0.000	1.603×10^{-5}	0.000	0.008
64QAM.TB.Retrans (X_{12})	0.016	0.002	0.000	-4.992×10^{-5}	0.000	0.000
QPSK.TB.Retrans (X_{13})	-0.025	0.002	0.000	-1.762×10^{-5}	0.000	0.000
16QAM.TB (X_{14})	0.004	0.000	0.000	1.695×10^{-6}	0.000	0.001
64QAM.TB (X_{15})	-0.005	0.000	0.000	-8.159×10^{-6}	0.000	0.000
QPSK.TB (X_{16})	0.003	0.000	0.000	1.049×10^{-5}	0.000	0.000

Another regression model defines the linear connection of the variable *Average User Downlink Throughput* and the observed set of the independent variables and can be written in the following form:

$$Y_2 = -11.194 - 0.941X_1 + 3.261X_2 + 0.001X_3 - 0.103X_4 + 0.202X_5 + 0.046X_6 + 1.031X_7 - 7.331 \cdot 10^{-5}X_8 + 7.704 \cdot 10^{-5}X_9 + 4.820 \cdot 10^{-5}X_{10} + 1.603 \cdot 10^{-5}X_{11} - 4.992 \cdot 10^{-5}X_{12} - 1.762 \cdot 10^{-5}X_{13} + 1.695 \cdot 10^{-6}X_{14} - 8.159 \cdot 10^{-6}X_{15} + 1.049 \cdot 10^{-5}X_{16} \tag{5}$$

the value of the determination coefficient of this model is $R^2 = 0.543$.

5.2. Multilayer Perceptron Model (MLP)

The proposed MLP model is a supervised machine learning model that, by training on a training data set, learns the function that performs mapping $R^m \rightarrow R^o$. In this case, the dimension of the input space is $m = 17$, and the output $o = 2$.

According to the algorithm given in Section 3, the first step in creating an MLP model in the IBM SPSS technology involves preprocessing the data, i.e., rescaling or reducing all input values to the same range. In this study, a standardization method was used according to which the variables are rescaled binary so that they have an arithmetic mean equal to 0 and a standard deviation equal to 1.

In order to select the best final model, a total of 30 variants of the MLP model were tested for prediction of the observed dependent variables with different combinations of ratios of the set sizes for training and testing (30%:70%, 40%:60%, 50%:50%, 60%:40%, 70%:30%), and different values of the following metaparameters:

(1) *Number of neurons in hidden layers.* The architecture of all tested model variants consists of one input layer with 17 neurons, one output layer with two neurons and two hidden layers between them, whose size varies in combinations and can be: 10,10; 20,20; 30,30.

(2) *The shape of the output function of neurons in hidden layers.* The IBM SPSS technology offers the possibility to select one of the two functions, which were examined in combinations with the first metaparameter of the model, namely hyperbolic tangent (*Hyp. Tg.*) and sigmoid function (*Sig.*). Each of the 30 variants of the MLP model with the above combinations were tested by multiple passes through the multilayer network architecture, i.e., triple consecutive trainings and tests, and the results are presented by the relative testing errors in Table 7 for both dependent variables. Based on the minimum value of the arithmetic mean of this relative accuracy criterion for the three measurements, the final model was selected.

According to [48], machine learning models are most often trained with 70% of the total data set, while the remaining 30% of the set is reserved for testing, i.e., to assess the accuracy and quality of the model. Additionally, in [49] the division of data in the mentioned ratio showed a good influence on the accuracy of the classifier. The random forest model presented in [50] showed the lowest mean standard error when dividing the data in the ratio of 70%:30%.

The selection of this ratio when creating predictive models is one of the inevitable steps that affects its performance. Based on the results of multiple tests of 30 different model variants in Table 7, it is concluded that the best performance is shown by the MLP model variant in which the division of the total data set of 64,301 input/output vector was performed in the ratio 70%:30%.

The average RE of the selected model for the output variable *Cell Downlink Average Throughput* is $RE = 0.104$ which means that the degree of accuracy is equal to 89.6%. For the variable *Average User Downlink Throughput*, the average relative error has the value $RE = 0.120$, i.e., the degree of accuracy is 88%.

Table 7 also shows that each hidden layer of the architecture of the selected model consists of 20 neurons, whose output function has the shape of the hyperbolic tangent $\tanh x$ (Figure 5). The output function of the two neurons in the output layer for the selected, but also for all tested variants of the MLP model, is linear or *Identity*.

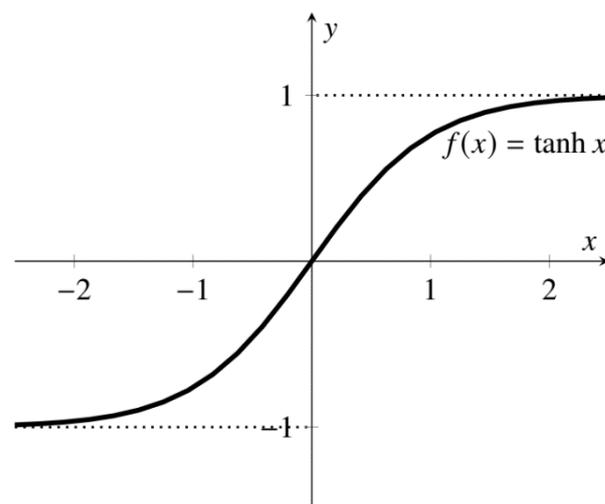


Figure 5. Output function of neurons shape of hyperbolic tangent $\tanh x$.

Table 7. Relative testing errors of the tested models.

The Ratio of Data Sets for Training and Testing	No of Neurons in Hidden Layers	Output Function of Neurons in Hidden Layers	Cell Downlink Average Throughput Measurements				Average User Downlink Throughput Measurements			
			1.	2.	3.	Average	1.	2.	3.	Average
30%:70%	10, 10	Hyp. Tg.	0.124	0.127	0.129	0.127	0.141	0.141	0.141	0.141
		Sig.	0.133	0.145	0.134	0.137	0.145	0.144	0.141	0.143
	20, 20	Hyp. Tg.	0.114	0.123	0.131	0.123	0.129	0.137	0.148	0.138
		Sig.	0.124	0.134	0.131	0.130	0.136	0.147	0.145	0.143
	30, 30	Hyp. Tg.	0.120	0.113	0.122	0.118	0.140	0.134	0.133	0.136
		Sig.	0.135	0.118	0.124	0.126	0.145	0.132	0.135	0.137
40%:60%	10, 10	Hyp. Tg.	0.121	0.125	0.147	0.131	0.136	0.137	0.157	0.143
		Sig.	0.127	0.148	0.133	0.136	0.141	0.145	0.148	0.145
	20, 20	Hyp. Tg.	0.114	0.129	0.119	0.121	0.131	0.150	0.136	0.139
		Sig.	0.123	0.115	0.119	0.119	0.141	0.131	0.132	0.135
	30, 30	Hyp. Tg.	0.117	0.111	0.111	0.113	0.140	0.131	0.129	0.133
		Sig.	0.129	0.117	0.119	0.122	0.145	0.137	0.133	0.138
50%:50%	10, 10	Hyp. Tg.	0.123	0.141	0.122	0.129	0.134	0.164	0.141	0.146
		Sig.	0.170	0.120	0.211	0.167	0.158	0.137	0.188	0.161
	20, 20	Hyp. Tg.	0.112	0.120	0.114	0.115	0.128	0.124	0.138	0.130
		Sig.	0.122	0.130	0.121	0.124	0.137	0.138	0.132	0.136
	30, 30	Hyp. Tg.	0.109	0.117	0.126	0.117	0.123	0.135	0.137	0.132
		Sig.	0.132	0.117	0.120	0.123	0.140	0.133	0.134	0.136
60%:40%	10, 10	Hyp. Tg.	0.118	0.115	0.147	0.127	0.138	0.135	0.160	0.144
		Sig.	0.127	0.123	0.128	0.126	0.137	0.144	0.140	0.140
	20, 20	Hyp. Tg.	0.105	0.118	0.111	0.111	0.136	0.140	0.133	0.136
		Sig.	0.120	0.112	0.125	0.119	0.139	0.130	0.137	0.135
	30, 30	Hyp. Tg.	0.114	0.115	0.115	0.115	0.129	0.133	0.128	0.130
		Sig.	0.116	0.122	0.114	0.117	0.135	0.138	0.130	0.134
70%:30%	10, 10	Hyp. Tg.	0.123	0.138	0.127	0.129	0.142	0.153	0.141	0.145
		Sig.	0.131	0.117	0.127	0.125	0.140	0.131	0.139	0.137
	20, 20	Hyp. Tg.	0.107	0.098	0.107	0.104	0.124	0.115	0.121	0.120
		Sig.	0.122	0.112	0.108	0.114	0.136	0.123	0.125	0.128
	30, 30	Hyp. Tg.	0.117	0.117	0.126	0.120	0.140	0.140	0.133	0.138
		Sig.	0.118	0.121	0.111	0.117	0.132	0.129	0.124	0.128

Table 8 shows the variants of the tested MLP models that have the lowest average values of testing error by individual ratios of the training data set and the testing data set. This table is a summary of Table 7. Based on Table 8, it can be concluded that as the percentage of training data increases, the average relative prediction error decreases, e.g., it is obvious that for a ratio of 50%:50% with Hyp.Tg. and with 20 neurons in each hidden layer, the model has an average RE = 0.115, so the degree of accuracy is 88.5%.

If this value is compared with RE = 0.104, i.e., with a degree of accuracy of 89.6%, as much as a model has with a ratio of 70%:30%, it is concluded that by reducing the ratio to 50%:50%, we lose only 1.1% of the performance accuracy of the model. This means that the model can perform prediction quite well even at a ratio of 50%:50%.

Table 8. Least average values of RE by examined ratios of training data set and test data set.

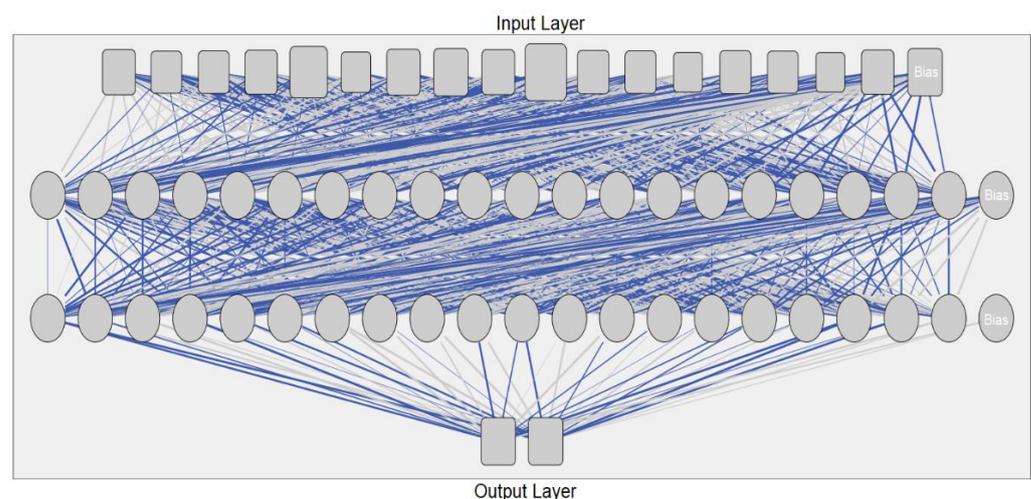
The Ratio of Data Sets for Training and Testing	No of Neurons in Hidden Layers	Output Function of Neurons in Hidden Layers	Cell Downlink Average Throughput Measurements				Average User Downlink Throughput Measurements			
			1.	2.	3.	Average	1.	2.	3.	Average
30%:70%	30, 30	Hyp. Tg.	0.120	0.113	0.122	0.118	0.140	0.134	0.133	0.136
40%:60%	30, 30	Hyp. Tg.	0.117	0.111	0.111	0.113	0.140	0.131	0.129	0.133
50%:50%	20, 20	Hyp. Tg.	0.112	0.120	0.114	0.115	0.128	0.124	0.138	0.130
60%:40%	20, 20	Hyp. Tg.	0.105	0.118	0.111	0.111				
70%:30%	30, 30	Hyp. Tg.					0.129	0.133	0.128	0.130
	20, 20	Hyp. Tg.	0.107	0.098	0.107	0.104	0.124	0.115	0.121	0.120

According to Table 7, the lowest individual values of relative testing errors among multiple iterations for the selected MLP model are as follows:

- For the variable *Cell Downlink Average Throughput* RE = 0.098 (degree of accuracy 90.2%);
- For the variable *Average User Downlink Throughput* RE = 0.115 (degree of accuracy 88.5%).

Figure 6 shows the two-layer architecture of the MLP model which achieved the above performance, where the synaptic connections, which are represented by dark blue, have weight factors less than zero, while the lighter color represents positive values of the weight factors.

The metaparameters of multilayer perceptron training do not change during the testing of different models. It is specified that the adjustment of the weighting factors is performed after processing each individual input/output vector from the training set (individual or *online* learning). The training algorithm is backpropagation that implements the gradient descent method. The *Initial Learning Rate* has a great influence on the convergence of the algorithm and its value is usually chosen from the range of 10^{-6} to 1. A higher learning rate accelerates the training process of the MLP model, but negatively affects its prediction performance. The initial value of the learning rate gradually decreases in each epoch, which represents one passage of a complete set of training data through the training algorithm, to a certain lower limit. The IBM SPSS technology specifies a *default* value of the initial learning rate of 0.4 while its lower limit is 0.001. The maximum number of epochs is determined automatically.

**Figure 6.** Multilayer architecture of the created model MLP.

The following part of this section provides detailed training and testing results for the selected final MLP model with the lowest individual RE values for the two dependent variables (RE = 0.098 for the *Cell Downlink Average Throughput* variable and RE = 0.115 for the *Average User Downlink Throughput* variable). Based on the results shown in Table 9,

generated by IBM SPSS technology, it can be seen that the relative error, i.e., the degree of accuracy of the model over the training data, for the variable *Cell Downlink Average Throughput* is 89.8% (RE = 0.102) and is slightly lower in relation to the degree of accuracy for the test set (the degree of accuracy is 90.2%, RE = 0.098), while for the variable *Average User Downlink Throughput*, the degree of accuracy over the training data is 88.6% (RE = 0.114), and for testing 88.5% (RE = 0.115). During the training process, the goal is to minimize the error function which, in the observed case, is represented by the *Sum of Squares Error* (SSE), and is calculated according to the equation:

$$SSE = \sum_{i=1}^N (RV_i - PV_i)^2 \tag{6}$$

where *RV* (*Real Value*) represents the actual value of the observed dependent variable, *PV* (*Predicted Value*) represents the prediction value of the observed dependent variable, and *N* is the number of all measurements. In the initial state, when the model is not yet trained, the value of the sum of the square error is significantly higher and is calculated by using the values of the arithmetic mean of that dependent variable as the prediction values of the observed dependent variable for all measurements. According to Table 9, the error function after model training has a value of 4882.801 and when testing the model, its value is smaller and amounts to 2089.438.

Two criteria for stopping the training process were set: (1) expiration of a time period of 15 min and (2) reaching the number of five consecutive iterations without reducing the error function. By meeting at least one of the above criteria, the training process is interrupted, and in the specific case of the selected MLP model, condition 2 is met, as can be seen in Table 9.

Table 9. The most important results of training and testing of the selected MLP model.

		Sum of squares error	4882.801
		Average overall relative error	0.108
Training	Relative error for scale dependents	Cell downlink average throughput	0.102
		Average user downlink throughput	0.114
		Stopping rule used	Five consecutive step(s) with no decrease in error
		Training time	0:01:00
		Sum of squares error	2089.438
		Average overall relative error	0.106
Testing	Relative error for scale dependents	Cell downlink average throughput	0.098
		Average user downlink throughput	0.115

Figure 7 graphically shows the best individual result from the set of 30 variants of the MLP model testing, in the form of *scatter plots*, i.e., prediction results for the two observed dependent variables. The abscissa shows the values from the test set, while the ordinate indicates the prediction values. Based on the obtained coordinates of the points and their arrangement, it is possible to visually evaluate the accuracy of the model. This way of presenting the prediction results is suitable for determining the coefficient of determination (R^2) which describes the percentage of variance of the target variable, which is explained by the model. For the dependent variable *Cell Downlink Average Throughput*, the best individual result of training and testing gives the value of the coefficient of determination $R^2 = 0.899$ while for the *Average User Downlink Throughput*, this value is $R^2 = 0.885$. The coefficient of determination is higher if the points are grouped around a line defined by the equation $y = x$, which is confirmed in Figure 7. If the R^2 values obtained in this way are compared with those calculated for the linear models, it can be concluded that MLP has a significantly higher accuracy of performance prediction or model quality because

- for cell downlink average throughput: $0.899 > 0.541$ and
- for average user downlink throughput: $0.885 > 0.543$.

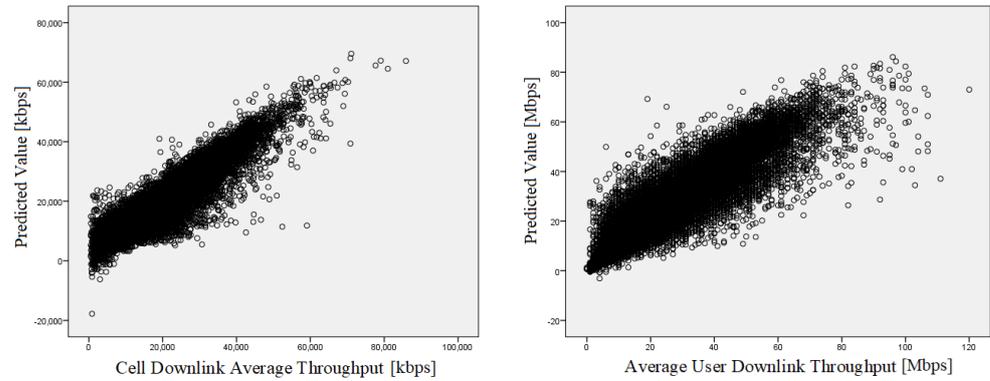


Figure 7. Scatter plots of the model testing results.

The *Chaddock* scale, presented in Table 10 [51], was used to qualitatively evaluate the coefficient of determination R^2 . It can be seen that the accuracy performance of the MLP model is in the range of 0.7–0.9 and qualifies as high. On the other hand, the accuracy of prediction of linear models based on the coefficients of determination can be assessed as *salient*. Comparing the prediction accuracy of the MLP model with the linear model, it is obvious that it has higher, i.e., high prediction accuracy, but it is also characterized by incomparably higher complexity compared to the linear model.

Table 10. Chaddock scale.

Determination Coefficient	Qualitative Characteristic
0.1–0.3	Weak
0.3–0.5	Moderate
0.5–0.7	Salient
0.7–0.9	High
0.9–0.99	Very high

As a quality parameter of the created predictive model, used *Residual* (r) is used which represents the difference between the real value (RV) and the prediction value (PV) of the observed variable: $r = RV - PV$. The arrangement of the residuals for the two output variables is shown in Figure 8. The model is of better quality and more accurate if it has residuals grouped around the horizontal zero line, which, in this case, is visually clearly visible.

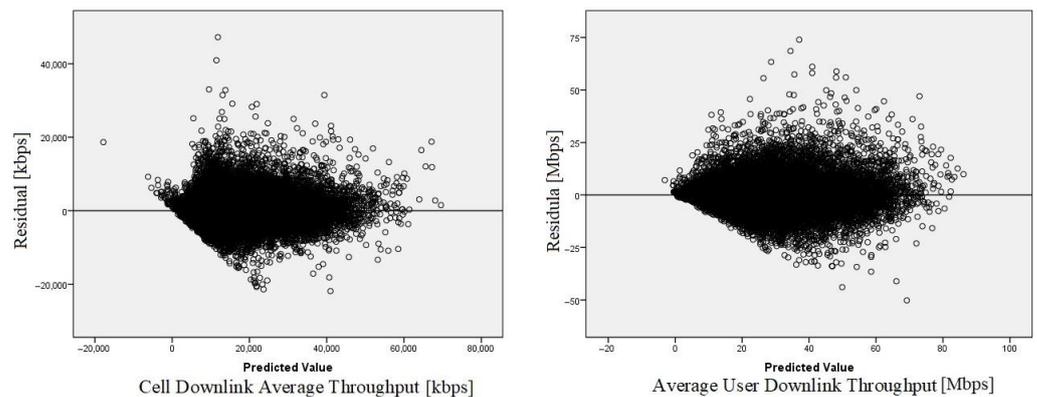


Figure 8. Residual distribution graph for two output variables in MLP prediction model.

One of the graphical results, generated by IBM SPSS technology, which represents the ranking of independent variables according to the *importance of influencing* the variability of the dependent variable in the model is shown in Figure 9. The measure showing how much the value created by the prediction changes for different values of the independent variable represents the *importance* of the independent variable expressed numerically at the lower position of the abscissa of the graph, while the *normalized importance*, obtained by dividing the importance with the largest individual value, is expressed as a percentage and presented at the upper position of the abscissa of the graph in Figure 9.

The interpretability of the model, as its ability to present the results to the researcher in an understandable degree, can be seen in Figure 9 and indicates that the variable *DL Cell Traffic Volume* has the greatest *importance* of influencing the changes in the values of the dependent variables. This conclusion is logical and expected considering that the throughput in the network, by definition, directly depends on the total realized traffic or the amount of data transmitted during the observed time. In contrast, the *Average CQI* variable is at the bottom of the rankings and has the least importance of influencing the changes in the value of the dependent variable. This is explained by the fact that MLP well models the nonlinear dependences of output on input. The observed variable is linearly related to the corresponding modulation and coding scheme of communication, and it is important for the size of the transport block when transmitting data through a channel which directly influences the throughput of specific classes of traffic.

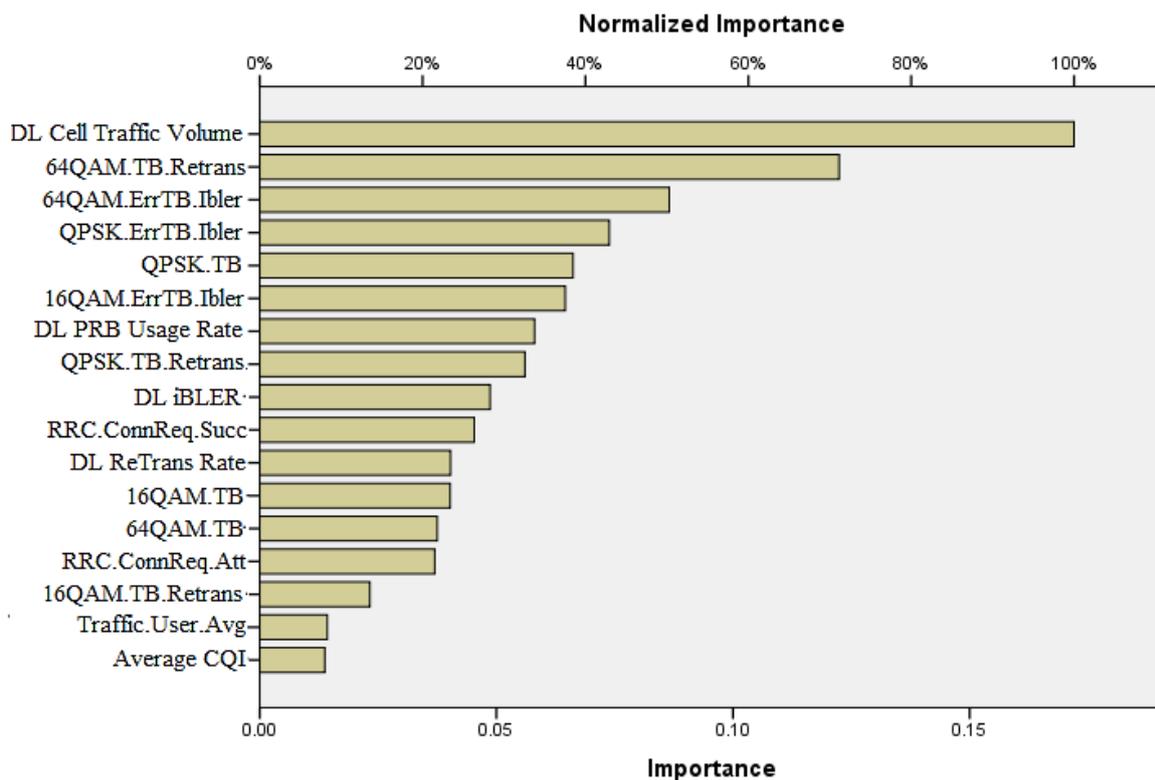


Figure 9. Graphical overview of the results of the influence of the importance of individual independent variables in the MLP prediction model.

Based on the fact that the creation of the model in this paper requires previously collected research data related to the network traffic and its parameters, it is concluded that, in this case a statistical approach is applied to the traffic modeling, which is otherwise based on *traces* (recording of combined traffic). Another possible approach is through an *emulator*, and, in some cases, a *hybrid* approach is possible. Since the variable *DL Cell Traffic Volume* means that it is the total traffic in the DL direction at the cell level, the created predictive model is a *model of combined traffic*. When models are created for each service, i.e.,

for each of the nine traffic classes shown in Table 5, this means that an approach based on *models of information sources* is applied, which are characterized by greater accuracy, but also greater complexity than the previous one. According to [22], the models of combined traffic are, in most cases, more suitable for application in the prediction of traffic in networks, and in [23] a good overview of the traffic models is presented.

The analysis of possible differences in the results of training and testing of different variants of the model in the three iterations shown was carried out by appropriate statistical tests. If the three measurements of the RE values are considered as three statistical groups, each tested variant of the MLP model represents one experimental unit over which the experiment is repeated three times. A parametric technique for testing statistical hypotheses about the equality of arithmetic means for three or more groups when it comes to repeated measurements is the analysis of variance or ANOVA of the repeated measurements. One of the basic preconditions for its application, in this case, is the normality of the distribution of the RE by the observed groups, i.e., the repeated measurements.

Table 11 shows the results of the *Kolmogorov–Smirnov* and *Shapiro–Wilk* normality tests, based on which it can be concluded that the distribution of the RE deviates from the normal distribution in all three groups for the variable *Cell Downlink Average Throughput* (Sig. < 0.05) [52], while only the first and second groups of the repeated relative error measurements for the *Average User Downlink Throughput* variable follow the normal distribution (Sig. > 0.05).

Table 11. Results of the normality tests of the RE by repeated measurements.

Dependent Variable	Measurement/Group	Kolmogorov–Smirnov			Shapiro–Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
<i>Cell Downlink Average Throughput</i>	1.	0.173	30	0.022	0.811	30	0.000
	2.	0.161	30	0.046	0.924	30	0.035
	3.	0.235	30	0.000	0.668	30	0.000
<i>Average User Downlink Throughput</i>	1.	0.129	30	0.200	0.944	30	0.117
	2.	0.122	30	0.200	0.964	30	0.392
	3.	0.226	30	0.000	0.816	30	0.000

Thus, ANOVA of the repeated measurements in this case cannot be used to determine potential statistical equations or differences between the mentioned groups. An appropriate nonparametric technique applied in case of nonfulfillment of the prerequisites for the analysis of variance is the *Friedman* test conducted in the IBM SPSS technology and whose results are shown in Table 12 for both dependent variables.

Table 12. Results of Friedman test for two dependent variables.

Cell Downlink Average Throughput		Average User Downlink Throughput	
N	30	N	30
Chi-Square	1.914	Chi-Square	1.270
df	2	df	2
Asymp. Sig.	0.384	Asymp. Sig.	0.530

Based on the value of *Asimp. Sig.* ($0.384 > 0.05$ and $0.530 > 0.05$), it can be concluded that there are no statistically significant differences in the relative errors, i.e., the prediction accuracy performance by the individual training and testing iterations of the 30 observed MLP models.

6. Conclusions

Current directions of development of cellular networks, which cover areas of different topological characteristics with a signal, are oriented towards providing permanent Internet access for mobile users both in rural areas and along roads. The selected case study in this

paper is the traffic segment M9J, in RS, BiH with access roads with an explanation of the high importance of this road and its connection with the key telecommunications network, M:tel BL. The starting idea is related to the fact that the predictive modeling, as a procedure of using known data in creating models for predicting future results, is significantly improved with the adaptive modeling of the network performance. The specific example refers to the potential application of the connected vehicle concept, for interactive vehicle-to-vehicle communication and with road infrastructure (vehicle-to-infrastructure). Additionally, the developed model can be an important reference for designing a network of wireless sensors placed along motorways that can communicate within the LTE network technology, which are part of the CPS. This means that the presented results enable the mobile operator to better analyze network performance for the needs of future planning and design of traffic in next generation networks.

The results of training and testing of 30 different variants of the MLP model, with different metaparameters showed that the final prediction model for the variable Cell Downlink Average Throughput has an average accuracy of 89.6%, while for the variable Average User Downlink Throughput the average accuracy is 88%. Additionally, observing the value of R^2 , it is concluded that in the case of the first dependent variable, the accuracy of the model is 1.4% higher than the accuracy of the model for the second dependent variable.

The classical predictive models, such as the linear regression ones, are characterized by training speed and simplicity. The research in this paper confirms that adaptive models based on machine learning, including MLP, show their key advantages in better modeling the nonlinear functional dependence of output on input. It means dependent on independent variables, especially when their number is higher. When accuracy is taken into account, the superiority of the MLP model over the MLR model can be expressed quantitatively based on R^2 . The best MLP model for Cell Downlink Average Throughput prediction has 35.8% higher accuracy compared to the same MLR model. For the Average User Downlink Throughput variable, the advantage of the MLP model over the MLR is higher by 34.2%. A special advantage of the developed MLP model in the IBM SPSS technology is given by a large number of tabular and graphical results that significantly facilitate the understanding of the relationship between variables. The obtained values of RE on the test data set, as well as the coefficients of determination, indicate high performance of the model, but also its possible improvement.

Compared to conceptually similar analyzed previous research results in the literature, the MLP model created in this paper has completely different metaparameters, uses a larger set of independent variables (17) as inputs, but also a significantly larger training data set with a larger number of training and testing set combinations. The selected final MLP model, based on indicators of training and testing with multiple passes through the network structure, as many as 30 variants of different combinations of the metaparameters in three iterations, significantly contributes to the validity and “reality” of the research results and its application in the network performance prediction. Table 1 presents an overview of the referenced previous research in the subject of this paper with highlighted novelties presented in this paper. The quality of the created model is further enhanced by the purposeful application in a specific geo-space with a reasoned analysis of the interactions of virtual and physical devices of the M:tel network and the M9J road. This additional dimension of adaptive modeling with ANN expands the criteria of functional, social and contextual adaptability of models with a biological dimension. It is because the enormous increase in the amount of transmission of different types of data (BigData) increased end-users’ interests in QoS and raised awareness of the importance of influence of the QoE to their multidimensional satisfaction.

The scope of this research can be expanded with a focus on finding methods to improve the performance of the created model by changing its parameters, metaparameters and architecture itself. That includes changing the number of layers and the number of artificial neurons in hidden deep network layers. Additionally, modeling delays and packet loss in mobile networks, but also creating KPIs models by individual traffic classes, is a

logical addition to the subject of future research by applying adaptive ANN models in traffic prediction in networks. Comparing the method presented in this paper with other competing methods such as comparison and interpretation methods for predictive control is one of the main directions of future research.

Author Contributions: Conceptualization, M.K.B. and M.S.; methodology, M.K.B., D.D. and M.S.; software, M.S.; validation, M.K.B. and D.D.; formal analysis, M.K.B. and M.S.; investigation, M.K.B., M.S. and D.D.; resources, M.K.B., Z.Ć., Z.M. and A.S.; data curation, M.K.B., M.S., Z.Ć., Z.M. and A.S.; writing—original draft preparation, M.K.B. and M.S.; writing—review and editing, M.K.B., M.S. and D.D.; visualization, M.S. and M.K.B.; supervision, M.K.B. and D.D.; project administration, Z.Ć., Z.M. and A.S.; financing acquisition, Z.Ć.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Faculty of Transport and Traffic Engineering Doboj University of East Sarajevo. This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Banjanin, K.M. *Komunikacioni Inženjering*; Univerzitet u Istočnom Sarajevu, Saobraćajno-Tehnički Fakultet Doboj: Doboj, Bosnia and Herzegovina, 2007; ISBN 978-99938-859-4-8.
- Walelgne, E.A.; Manner, J.; Bajpai, V.; Ott, J. Analyzing throughput and stability in cellular networks. In Proceedings of the NOMS 2018-2018 IEEE/IFIP Network Operations and Management Symposium, Taipei, Taiwan, 23–27 April 2018; pp. 1–9.
- Seetharam, A.; Walker, P. An empirical characterization of cellular network performance. In Proceedings of the 2016 International Conference on Computing, Networking and Communications (ICNC), Kauai, HI, USA, 15–18 February 2016; pp. 1–5.
- Casas, P.; Sackl, A.; Schatz, R.; Janowski, L.; Turk, J.; Irmer, R. On the quest for new KPIs in mobile networks: The impact of throughput fluctuations on QoE. In Proceedings of the 2015 IEEE International Conference on Communication Workshop (ICCW), London, UK, 8–12 June 2015; pp. 1705–1710. [[CrossRef](#)]
- Huang, J.; Xu, Q.; Tiwana, B.; Mao, Z.M.; Zhang, M.; Bahl, P. Anatomizing application performance differences on smartphones. In Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, San Francisco, CA, USA, 15–18 June 2010; pp. 165–178.
- Sonntag, S.; Schulte, L.; Manner, J. Mobile network measurements-It's not all about signal strength. In Proceedings of the 2013 IEEE Wireless Communications and Networking Conference (WCNC), Shanghai, China, 7–10 April 2013; pp. 4624–4629. [[CrossRef](#)]
- Abed, G.A.; Ismail, M.; Jumari, K. Modeling and performance evaluation of LTE networks with different TCP variants. *Proc. World Acad. Sci. Eng. Tech.* **2011**. [[CrossRef](#)]
- Bui, N.; Michelinakis, F.; Widmer, J. A model for throughput prediction for mobile users. In Proceedings of the European Wireless 2014; 20th European Wireless Conference, Barcelona, Spain, 14–16 May 2014; pp. 1–6.
- Samba, A.; Busnel, Y.; Blanc, A.; Dooze, P.; Simon, G. Instantaneous throughput prediction in cellular networks: Which information is needed? In Proceedings of the 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM), Lisbon, Portugal, 8–12 May 2017; pp. 624–627.
- Liu, Y.; Lee, J.Y. An empirical study of throughput prediction in mobile data networks. In Proceedings of the 2015 IEEE Global Communications Conference (GLOBECOM), San Diego, CA, USA, 6–10 December 2015; pp. 1–6. [[CrossRef](#)]
- Popović, M.; Drajić, D.; Svoboda, P.; Nikaein, N.; Krčo, S.; Laner, M. Latency analysis for M2M and Online Gaming traffic in an HSPA network. *Adhoc Sens. Wirel. Netw.* **2016**, *31*, 259–277.
- Popović, M.; Drajić, D.; Krčo, S. Evaluation of the UTRAN (HSPA) performance in different configurations in the presence of M2M and online gaming traffic. *Trans. Emerg. Telecommun. Technol.* **2016**, *26*, 800–812. [[CrossRef](#)]
- Isak-Zatega, S.; Lipovac, A.; Lipovac, V. Logistic regression based in-service assessment of mobile web browsing service quality acceptability. *EURASIP J. Wirel. Commun. Netw.* **2020**, 1–21. [[CrossRef](#)]
- Liang, D.; Zhang, J.; Jiang, S.; Zhang, X.; Wu, J.; Sun, Q. Mobile Traffic Prediction Based on Densely Connected CNN for Cellular Networks in Highway Scenarios. In Proceedings of the 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP), Xi'an, China, 23–25 October 2019; pp. 1–5.
- Rehman, T.; Baig, M.A.I.; Ahmad, A. LTE downlink throughput modeling using neural networks. In Proceedings of the 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), New York, NY, USA, 19–21 October 2017; pp. 265–270.
- 3GPP TS 36.213. *Physical Layer Procedures*, v12.5.0; 3rd Generation Partnership Project: Sophia Antipolis, France, 2015.
- 3GPP TS 36.314. *Physical Layer Procedures*, v8.1.0; 3rd Generation Partnership Project: Sophia Antipolis, France, 2009.

18. 3GPP TS 29.155. *Representational State Transfer over St Reference Point*, v13.0.0; 3rd Generation Partnership Project: Sophia Antipolis, France, 2016.
19. 3GPP TS 25.322. *Radio Link Control (RLC) Protocol Specification*, v7.10.0; 3rd Generation Partnership Project: Sophia Antipolis, France, 2009.
20. 3GPP TS 36.314. *Layer 2 Measurements*, v10.2.0; 3rd Generation Partnership Project: Sophia Antipolis, France, 2011.
21. 3GPP TS 36.824. *LTE Coverage Enhancements*, v11.0.0; 3rd Generation Partnership Project: Sophia Antipolis, France, 2012.
22. Drajić, D. *Uvod u M2M (Machine-to-Machine) Komunikacije*; Akademski misao, Univerzitet u Beogradu—Elektrotehnički Fakultet: Beograd, Serbia, 2016; ISBN 978-86-7466-591-6.
23. Nikaein, N.; Drajić, D.; Svoboda, P.; Popovic, M.; Krco, S. *Chapter_03_M2M Traffic and Models*; Markus Laner, Machine-To-Machine Communications—Architectures, Technology, Standards, and Applications; Vojislav, B., Mišić, J.M., Eds.; Publisher CRC Press: Boca Raton, FL, USA, 2014; ISBN 978-1-46-656123-6. Available online: <http://www.crcpress.com/product/isbn/9781466561236> (accessed on 20 March 2021).
24. Putnik, G.; Ferreira, L.; Lopes, N.; Putnik, Z. What is a Cyber-Physical System: Definitions and Models Spectrum. *FME Trans.* **2019**, *47*, 4. [[CrossRef](#)]
25. Duch, W.; Jankowski, N. Survey of neural transfer functions. *Neural Comput. Surv.* **1999**, *2*, 163–212.
26. Elizondo, D. The linear separability problem: Some testing methods. *IEEE Trans. Neural Netw.* **2006**, *17*, 330–344. [[CrossRef](#)] [[PubMed](#)]
27. Yanhong Ma; Shuanglei, F. *Large Scale Wind Power Grid Integration: Technological and Regulatory Issues*; Ningbo, W., Chongqing, K., Dongming, R., Eds.; Elsevier: Amsterdam, The Netherlands, 2016; ISBN 978-0-12-849895-8. [[CrossRef](#)]
28. Abirami, S.; Chitra, P. Energy-efficient edge based real-time healthcare support system. In *Advances in Computers*; Elsevier: Amsterdam, The Netherlands, 2020; Volume 117, pp. 339–368.
29. Madan, M.G. *Soft Computing and Intelligent Systems Theory and Applications*; Sinha, N., Ed.; Academic Press: San Diego CA, USA, 1999.
30. Lonkar, S.A.; Reddy, K.V. Analysis of audio and video quality of voice over LTE (VoLTE) call. *Int. J. Inf. Technol.* **2020**, 1–14. [[CrossRef](#)]
31. Pavlović, Z.; Banjanin, M.; Vukmirović, J.; Vukmirović, D. Contactless ICT transaction model of the urban transport service. *Transport* **2020**, *35*, 5–500. [[CrossRef](#)]
32. Stjepanovic, A.; Banjanin, M. Distributed Multimedia Information System for Traffic Monitoring and Managing. In *Intelligent Data Engineering and Automated Learning—IDEAL 2014*; Corchado, E., Lozano, J.A., Quintián, H., Yin, H., Eds.; Lecture Notes in Computer Science, (Referred in WoS); Springer: Cham, Switzerland, 2014; Volume 8669, pp. 475–483. [[CrossRef](#)]
33. Banjanin, M.; Stojčić, M.; Stjepanović, A. *Analiza Algoritama Kodovanja Multimedijalnih Podataka*; Univerzitet u Istočnom Sarajevu Filozofski Fakultet Pale, časopis “Radovi” Filozofski Fakultet UIS, Filozofske i Prirodno-Matematičke Nauke: Pale, Bosnia and Herzegovina, 2012; Volume 2, pp. 179–192, ISSN 1512-5858. [[CrossRef](#)]
34. Stojčić, M.; Stjepanović, A.; Kostadinović, M.; Kuzmić, G.; Banjanin, M. Adaptive Neuro-Fuzzy Model for Traffic Signs Recognition. In Proceedings of the IEEE Xplore 2020 19th International Symposium INFOTEH-JAHORINA, East Sarajevo, Bosnia and Herzegovina, 18–20 March 2020. [[CrossRef](#)]
35. Stojčić, M.; Banjanin, M.; Jovanović, Ž.; Božičković, S. Adaptivni neuro-fazi model za predikciju nivoa snage Wi-Fi signala u zatvorenom prostoru. In Proceedings of the Zbornik Radova sa VII Međunarodnog Simpozijuma NOVI HORIZONTI 2019, Dobo, Bosnia and Herzegovina, 29–30 November 2019; pp. 568–576, ISBN 978-99955-36-79-4.
36. Sadeghi, A.; Luvisotto, M.; Lahouti, F.; Zorzi, M. Analysis of Statistical QoS in Half Duplex and Full Duplex Dense Heterogeneous Cellular Networks. *arXiv* **2017**, arXiv:1708.06215.
37. Liberg, O.; Sundberg, M.; Wang, E.; Bergman, J.; Sachs, J.; Wikström, G. *Cellular Internet of Things: From Massive Deployments to Critical 5G Applications*; Academic Press: Cambridge, MA, USA, 2019.
38. Abbas, S.S.A.; Thiruvengadam, S.J.; Punitha, M. Realization of PDSCH transmitter and receiver architecture for 3GPP-LTE advanced. In Proceedings of the 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 23–25 March 2016; pp. 1–6.
39. Loshakov, V.A.; Al-Janabi, H.D.; Al-Zayadi, H.K. Adaptive control signal parameters in LTE technology with MIMO. *J. Telecommun. Probl.* **2012**, *2*, 78–90.
40. Perez, A. *Implementing IP and Ethernet on the 4G Mobile Network*; Elsevier: Amsterdam, The Netherlands, 2017; ISBN 978-1-78548-238-0.
41. M:Tel Databanja Luka (Data Obtained from Vendor). M:Tel Company Banja Luka. Available online: <https://mtel.ba/> (accessed on 20 March 2021).
42. Wang, D.; Soni, R.; Chen, P.; Rao, A. Video telephony over downlink LTE systems with/without QoS provisioning. In Proceedings of the 34th IEEE Sarnoff symposium, Princeton, NJ, USA, 3–4 May 2011; pp. 1–5.
43. Rommer, S.; Hedman, P.; Olsson, M.; Frid, L.; Sultana, S.; Mulligan, C. *5G Core Networks*; Academic Press: Cambridge, MA, USA, 2019; ISBN 9780081030097.
44. Olsson, M.; Mulligan, C.; Rommer, S.; Sultana, S.; Frid, L. *SAE and the Evolved Packet Core: Driving the Mobile Broadband Revolution*; Academic Press: Cambridge, MA, USA, 2009.

45. Tarn, W.H.; Chen, J.; Lee, C.L. Fairness consideration of scheduling for real-time services in 4G systems. *Math. Probl. Eng.* **2015**. [[CrossRef](#)]
46. Obaidat, M.; Zarai, F.; Nicopolitidis, P. *Modeling and Simulation of Computer Networks and Systems*; Morgan Kaufmann: Waltham, MA, USA, 2015; ISBN 978-0-12-800887-4. [[CrossRef](#)]
47. OpenSignal Methodology Overview: How OpenSignal Measures Mobile Network Experience. Available online: https://www.opensignal.com/sites/opensignal-com/files/opensignal_methodology_overview_feb_2021.pdf (accessed on 29 January 2021).
48. Raschka, S. Model evaluation, model selection, and algorithm selection in machine learning. *arXiv* **2018**, arXiv:1811.12808.
49. Dahiya, S.; Handa, S.S.; Singh, N.P. Credit scoring using ensemble of various classifiers on reduced feature set. *Industrija* **2015**, *43*. [[CrossRef](#)]
50. Adelabu, S.; Mutanga, O.; Adam, E. Testing the reliability and stability of the internal accuracy assessment of random forest for classifying tree defoliation levels using different validation methods. *Geocarto Int.* **2015**, *30*, 810–821. [[CrossRef](#)]
51. Sobolev, I.; Babichenko, S. Application of the wavelet transform for feature extraction in the analysis of hyperspectral laser-induced fluorescence data. *Int. J. Remote Sens.* **2013**, *34*, 7218–7235. [[CrossRef](#)]
52. Murali, G.; Gayathri, R.; Ramkumar, V.R.; Karthikeyan, K. Two statistical scrutinize of impact strength and strength reliability of steel fibre-reinforced concrete. *KSCE J. Civ. Eng.* **2018**, *22*, 257–269. [[CrossRef](#)]