

Article

Studying Spatial Unevenness of Transport Demand in Cities Using Machine Learning Methods

Denis Chainikov ^{1,*}, Dmitrii Zakharov ¹, Evgeniy Kozin ² and Anatoly Pistsov ¹

¹ Department of Road Transport Operation, Transport Institute, Industrial University of Tyumen, 625000 Tyumen, Russia; zaharovda@tyuiu.ru (D.Z.); pistsovav@tyuiu.ru (A.P.)

² Department of Service of Vehicles and Technological Machines, Transport Institute, Industrial University of Tyumen, 625000 Tyumen, Russia; kozines@tyuiu.ru

* Correspondence: chainikovda@tyuiu.ru; Tel.: +7-905-8214301

Abstract: The article discusses the issues of spatial unevenness of transport demand in the city by various transport modes. It describes the creation of models using an artificial neural network to estimate the travel time and share by private and public transport in a large city that does not have off-street transport. The city transport macromodel in PTV Visum (V.18) was used as a data source, from which data were obtained on 50 basic parameters taken into account in the specialized software during the development of the transport model. In total, 50 factors act as independent variables that do not have linear relationships with each other and with the dependent variable, which significantly complicates the use of other models. These models allow assessing the influence degree of the most important factors. Further, the article shows the uneven spatial distribution of the shares of trips by private and public transport across the areas of a city. Using machine learning methods, the transport areas of Tyumen were clustered into nine classes belonging to the central sector, where the share of public transport is significantly higher than at the city border. The dependence of the trip share by cars and shuttle buses on the average travel time and distance by private and public transport for each class of transport areas has been established. The research results can be used when creating new transport areas in the city macromodel and when adjusting transport planning documents. The methods used for analyzing big data on the operation of the transport complex can be implemented in the digital twin of the city and the Intelligent Transport System.

Keywords: transport mobility; private and public transport; population mobility structure; artificial neural network; machine learning; road network



Citation: Chainikov, D.; Zakharov, D.; Kozin, E.; Pistsov, A. Studying Spatial Unevenness of Transport Demand in Cities Using Machine Learning Methods. *Appl. Sci.* **2024**, *14*, 3220. <https://doi.org/10.3390/app14083220>

Academic Editors: Tingsong Wang, Xu Guan and Kai Wang

Received: 12 February 2024

Revised: 8 April 2024

Accepted: 9 April 2024

Published: 11 April 2024



Copyright: © 2024 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

In the 21st century, the process of urbanization and suburbanization continues in many countries. With an increase in the population in cities, their residents receive services and improvements in the quality of some services and new advantages, but also problems in ensuring sustainable mobility for citizens [1]. Moreover, the COVID-19 pandemic has changed the structure of population transport mobility, reducing the demand for public transport [2]. The decline in demand for public transport affected the sustainability of the transport system. The authors of the study [3] assessed the sustainability of the city's transport system during a pandemic through the balance of transport demand and supply. The authors concluded that resilience varied over time and fluctuated between high and moderate levels. Almost 90% of the time, transport operators successfully coped with balancing between demand and supply of transport services.

For cities where the population is significantly changing (the area or the transport behavior model), it is very important to predict the possible change in the parameters of the transport system and determine the most effective measures for ensuring sustainable mobility. Such a tool is transport modeling using specialized software. Macroscopic

transport models are also used to forecast transport demand in new areas and determine plans for the development of road transport infrastructure in transport planning documents. In recent years, the topic of simulated transport modeling has become increasingly more interesting in the global scientific community [4]. When creating new built-up areas in a transport model, it is important to predict the number of trips for various purposes and the structure of residents' mobility in this area. After receiving a forecast assessment of transport demand, it is possible to determine an effective traffic management scheme for sections of the road network located inside and on the border of a new transport area [5].

When carrying out renovation or implementing integrated development projects for territories, the category of land plots in the city master plan changes and the development areas with private or mid-rise residential buildings are replaced with multi-story ones [6]. At the same time, the population density and transport mobility in the development area significantly increase. At the border of transport areas, in places where local driveways adjoin the supporting road network, traffic congestion occurs, and public transport operates with overload; i.e., the quality of transport services for the population declines. For an informed decision on infrastructure development, it is necessary to determine the correspondence matrix in the transport macromodel and the structure of population mobility. At the same time, due to the lack of residents in the new area, it is impossible to conduct sociological surveys, and there is a need to use other methods for forecasting and assessing the parameters of population mobility.

In the work [7], Zedgenizov A.V. notes that transport demand in suburban areas, which in the case study of Irkutsk include gardening partnerships and cottage villages, is influenced by the area of the territory. At the same time, the share of trips by private cars varies from 0.74 to 0.94. When clustering, all the main transport areas of Irkutsk are divided into six classes: residential; industrial; central part and business center; and areas with a high, moderate, and low share of housing area. In the work, the author established an influence pattern of the area's distance from the city center on the specific correspondence generation, which is described by a linear equation. In this case, it is necessary to take into account the seasonal unevenness of transport demand in relation to gardening partnerships in comparison with the areas of private residential development. At the same time, the number of residents in the city affects the volume of passenger transportation by public transport.

Transport behavior and mobility structure depend on the level of development of public transport and the road network [8]. At the same time, traffic routes and stopping points are connected to transport areas of the city and affect the level of development of public transport infrastructure in the city areas. The share of trips by private and public transport (PT) is a formal assessment of the transport behavior of the population. A number of works note the need to consider the concept of transport behavior from the point of view of psychology, sociology, and urban studies. An important element of urbanism is spatial settlement and the level of transport services for the population in city districts.

Research in the field of sustainable mobility has become actively widespread since the digitalization of the transport complex began and there emerged technical means that record the parameters of the urban transport system in real time or with a short time delay. The values of these parameters are necessary for planning the parameters of the city's transport infrastructure. For example, in [9], the parameters of transport infrastructure were used when planning mixed land use. Moreover, the number of studies of this kind aimed at socially sustainable transport is only growing [10]. To determine the parameters of the urban transport system, simulation modeling can be used on specialized software that requires the macromodel itself, data, and a lot of time and computational resources. The appearance on the city streets of an increasing number of detectors [11], video cameras, electronic methods of paying for travel on PT, parking, and rental vehicles (car sharing cars, bicycles, electric scooters) made it possible to generate big data and, through its processing, reduce the error and increase the adequacy of the patterns and models established by

researchers, as well as increase the accuracy of forecasting when carrying out simulation modeling of the operation of transport systems.

To process big data on the operation of urban transport systems when studying sustainable active mobility such as Mobility-on-Demand systems [12,13], bike sharing [14], and others [15], machine learning methods, including neural network technologies, are used.

The work [16] describes how deep learning models and neural networks are created to predict the structure of population mobility, i.e., the share of trips by private and public transport. The authors considered seven different models with varying accuracy. The model with an artificial neural network (ANN) turned out to be the most accurate.

The work [17] describes the use of ANNs for processing data on the actual transport mobility of the population by processing the GPS signal to further determine the mode of transport. The model determines the transportation mode by analyzing the trajectories and speeds obtained from the tracker. The authors study the effectiveness of different deep learning models and note the greatest efficiency of the ANN-based model.

Deep learning and ANNs are used to predict trips by taxi and bicycles using GPS, taking into account holidays, weather conditions, ambient temperature, and travel speed [18].

Computer vision using an ANN is used to determine the flow density from a video camera [19] with the purpose of further taking into account data on the traffic flow rate for operational traffic management, as well as assessing pollutant emissions from vehicle exhaust gases [20,21].

The authors of the work use a neural network to predict the travel purposes (destination) of passengers in taxi cars using GPS data, taking into account the travel request by time during the day [22]. This allows one to more accurately plan the work of taxi cars and select the optimal orders for each taxi car according to the criterion “length of the route to the passenger’s pick-up point”, which will increase the efficiency of transportation and reduce the possible waiting time for passengers.

ANNs are actively used to determine the optimal operating modes of traffic lights in traffic control. This is especially true during the transition from local adaptive control at one intersection to network adaptive control in a large area of the city and in the future when creating a top-level ITS. In this case, each traffic light unit at the intersection is a separate agent and is controlled separately in the system, taking into account the operating parameters of other agents [23].

An urgent task for municipalities is to predict the number of bicycle trips in order to determine the streets on which it is necessary to create infrastructure for cyclists, namely bicycle paths, parking areas, and special settings for traffic lights, and for businesses, the locations of bicycle sharing stations. The authors of the work use an ANN to forecast the demand for bicycle rentals and forecast the number of trips in space and time [24].

The purpose of this study is to establish a pattern of changes in the share of trips by private and PT, taking into account the spatial location of the transport area in a large city with a population of 800–900 thousand people and at a distance from the city center. To implement the model of the influence of factors, the method of artificial neural networks was chosen [25–27].

An analysis of previous papers showed that data related to the “digital shadow” of the transport system can be used to develop measures to ensure sustainable urban mobility and can be analyzed using machine learning methods and artificial neural networks.

2. Materials and Methods

The city transport macromodel consists of transport demand and transport supply models. The main element of the transport supply model is the transport area. It is the area in which the number of residents living in this territory is recorded, as well as the number of jobs in various sectors of the economy. Based on these data, the most important element for modeling, the correspondence matrix, is formed.

Transport areas are elementary units of the spatial structure of the modeling area.

The share of trips of city residents by various modes of transport is a parameter for assessing the transport behavior model. Increasing the share of trips by public transport and reducing the number of trips by private cars is a target function in municipal programs and projects to ensure sustainable mobility with a defined numerical value or mandatory conditions for achieving the goals of the project. For example, in Tyumen, in the transport planning document for 2040, the share of trips by PT should be increased from 0.41 to 0.55. To achieve the specified values, it is necessary to know what the share of trips by a mode of transport depends on and what factors have the greatest impact on this indicator.

The level of development of the road infrastructure depends on the number of exits from subdistricts to the supporting road network, traffic density, coefficient of non-straightness of traffic routes, etc. The level of development of the transport infrastructure depends on the density of the route network, the carrying capacity of sections of the road network, the connectivity of planned areas of the city with private and PT routes, and the number of top routes at stopping points, and it differs for the central and peripheral areas of the city.

Factors influencing the model of transport behavior are determined by the spatial characteristics of the city, the level of development of the road and route network, and the quality of the PT operation and include the area and shape of the city, the distance of the source area from the target area, the length of the private cars and PT traffic routes, trip time during peak hours and inter-peak times, and the number of transfers when traveling by PT.

This study examines the impact of travel time on the total number and share of trips taken by private or public transport in a large city without off-street transport.

The paper considers a hypothesis that the model of transport behavior and the structure of population mobility depend on the location of transport areas relative to the city center and that as travel time increases, the share of trips by personal transport increases, and the share of trips by public transport decreases.

The influence of the average travel time from the PT source area on the share of trips by public transport is determined by Equation (1):

$$\Delta_{PT} = \frac{1}{1 + e^{-(a+bT_{jPT})}}, \quad (1)$$

where Δ_{PT} is the share of trips by public transport; Δ_0 is the maximum value of the share of PT trips at the minimum value of T_{jPT} ; T_{jPT} is the average travel time of the PT from the j -th transport area source (target) of travel, min; a, b are model parameters.

The form of Equation (1) is determined by taking into account the use of a multinomial logit model to assess the probability of a user choosing a transportation mode. It is assumed that with minimal average values of PT travel time for the j -th transport area source of travel, the change in the share of PT trips is small. As the PT travel time value leaves the "comfort zone" for passengers, their desire to reduce time loss by using a personal car increases. With a significant increase in the travel time of the PT, the share of its use changes slightly and is determined by the group of people who do not have a personal car or the ability to use it.

The non-linear type of model 1 is determined by the following condition: any car driver can become a passenger of PT, but not any passenger of PT can become a car driver due to the lack of a personal car. In this case, two components are taken into account:

- The cost of travel by PT is significantly lower than by a private car.
- Within the city, with an increase in the route length and the possibility of free transfers on PT, the cost of traveling by PT does not change, while for private cars, it increases in direct proportion to the length of the route.

The influence of the average travel time by private cars from the source area on the share of trips by private transport is described by a linear equation:

$$\Delta_{\text{priv}} = \Delta_0 + S \times T_{j_{\text{priv}}}, \quad (2)$$

where Δ_{priv} is the share of trips by private transport; Δ_0 is the minimum value of the share of movements by private transport at the minimum value of $T_{j_{\text{priv}}}$; $T_{j_{\text{priv}}}$ is the average travel time by private transport from the j -th transport area source (target) of trip, min; S is a parameter of sensitivity to changes in trip time.

The difference in the type of mathematical models 1 and 2 is due to the possibility of private car users switching to traveling on foot or using bicycles and personal mobility devices with a short route length and travel time. When conducting research on the spatial unevenness of the transport behavior of residents and the share of trips by private or public transport, it is possible to use a city macromodel at the stage of obtaining initial data.

2.1. Description of Initial Datasets

A transport macromodel of a large city with a population of 800–850 thousand people can consist of 400 transport areas. This number of areas forms a correspondence matrix of 159,600 correspondence values between areas. With a high density of the road network and the presence of alternative traffic routes, the number of route options increases several-fold. On average, the transport macromodel of the city considers 4–5 transport systems and 8 travel goals, of which the main ones are work-related and educational goals. Dividing the correspondence matrix by purpose and transport means significantly increases the amount of data to be processed and analyzed. All this significantly complicates analyzing the results of modeling and assessing the parameters of the urban transport system. The PTV Visum (V.18) macromodeling program allows taking into account about 50 basic variables that affect the travel time, as well as creating additional attributes in the transport model.

All values of the variables of the model are quantitative indicators that vary over a wide range. Determining the travel time and the share of trips by modes of transport in the city macromodel requires a long time because of the functionality of the PTV Visum program and the availability of this software. To increase productivity in assessing the studied parameters and the possibility of conducting research without using expensive software, it was decided to create a neural network model for calculating the travel time.

The process of creating a neural network model within the framework of this study consisted of the following steps: obtaining and describing data; data loading; preparing a dataset (eliminating columns, determining independent variables X , dependent variable y , splitting into training and test sets, data standardization); creation of neural network architecture; search for optimal parameters of the neural network; training; prediction of y values; assessment of model accuracy; saving the model and its further use with the help of a computer program [28]. A mechanism was chosen to implement a fully connected neural network (an ANN model) with hidden layers for regression problems [29,30].

Using the data from the transport macromodel of the city created in the PTV Visum program, 3 datasets were obtained; they were used in the research with different goals and at different stages. The used macromodel of the city is shown in Figure 1.

The research was carried out in 3 stages.

In the first stage, the system evaluates the travel time (correspondence time) for all possible correspondence matrix variants. In this stage, the travel time is change-dependent and is determined by a large number of factors.

In the second stage, the shares of the use of travel modes and transport modes were assessed for each travel option. In this stage, the travel time will become the main factor affecting the share of trips by private and PT. Due to the large number of parameters evaluated at the first and second stages, a decision was made to use a neural network model to reduce the complexity of calculations.

In the third stage, clustering of city areas was carried out according to the criterion of the share of trips by private and PT. These two transport methods were chosen because they have the largest share in citizen mobility compared to pedestrian traffic or the use of cycling and personal mobility devices.

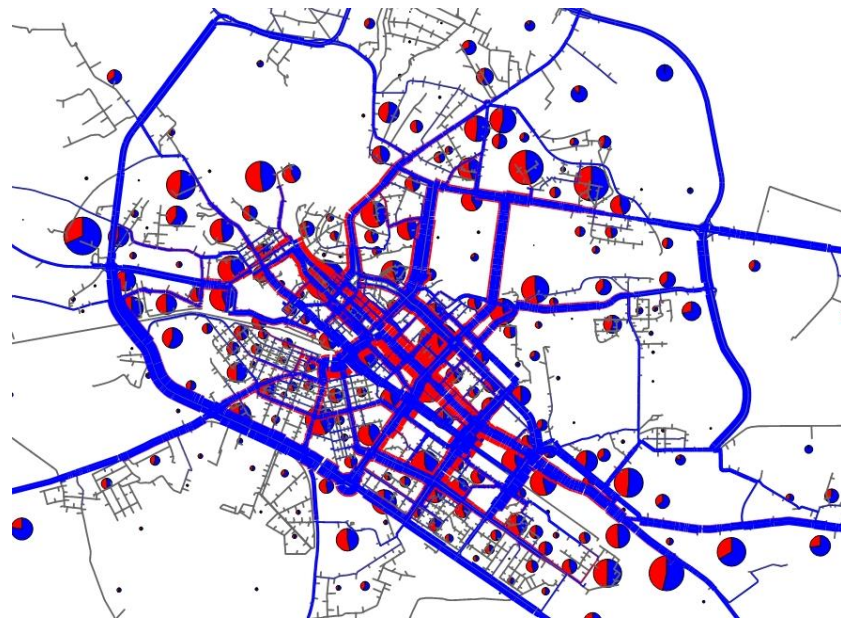


Figure 1. Macromodel of traffic flows in the city of Tyumen, Russia.

The transition from estimating the share of trips by correspondence to shares by area is due to several reasons:

- One road section can duplicate a large number of trips by private and PT, as well as public transport routes. In this case, there may be different effects and efficiencies between transport systems and individual trips within a single system for reducing time loss in traffic.
- The use of measures to reduce the share of trips by private cars has a greater effect when implementing measures that promote the use of PT and eliminating the reasons that restrain it. It is more efficient to eliminate data from large, systemic solutions, primarily for large areas of the city, than from separate, private solutions on sections of the road network.

After the third stage, the determination of measures for ensuring sustainable mobility requires simulation modeling and an assessment of the mobility structure and operating parameters of the transport system. If the level of these indicators does not match the target value, it is necessary to proceed to the fourth stage. This stage consists of clustering correspondence on separate transport systems (private and PT) and identifying the most problematic public transport routes and sections of the road network.

The first dataset is called “Traffic” and reflects the results of modeling the parameters of traffic between different areas of the city, taking into account the characteristics of these areas.

As a dependent variable (y), the variable t_a was considered—the travel time in a loaded network, s . Factors in the amount of 32 were considered as independent variables (X), among which we can single out the distance of the source area from the city center, the average number of lanes on the road segments, the number of traffic lights on the way, the total number of maneuvers performed, the total delay time at intersections, and the like. The choice of 32 parameters from the city transport macromodel related to the base parameters is due to their total influence on the attribute of the PTV Visum program resistance, which ultimately determines the travel time.

The dataset contains 139,341 rows (139,340 observations) and 33 columns corresponding to the number of dependent and independent variables.

As part of the second set of data, “Structure of mobility”, the influence of factors on the structure of the mobility of the city population was considered.

To plan the development of the urban transport network, among others, the indicator of the mobility structure is used, which expresses the percentage of types of transport used by the inhabitants of the area. As a rule, to determine the structure of mobility, simulation is performed using specialized software that requires significant time and computational resources. However, the model of the influence of factors on the structure of the mobility of the population of the city area has not yet been determined.

The resulting model will make it possible to predict the structure of the mobility of the urban population depending on a number of initial factors, among which we can distinguish the travel time, taking into account the load of the road network, the distance of the area of correspondence formation to the city center, the number of traffic light objects along the route of road transport, etc. In total, 50 factors act as dependent variables that do not have linear relationships with each other and with the independent variable, which significantly complicates the use of other regression models.

As the desired value (dependent variable y), the mobility structure of the city population is determined, which is expressed as a percentage of the following values: the share of trips by private transport ($car_%$); share of trips by public transport $bus_%$; share of trips by bike $bike_%$; share of pedestrian movements $ped_%$. The transposed dataset (first three rows) is presented in an abbreviated form in Table 1.

Table 1. Output of the “Structure of mobility” dataset in an abbreviated and transposed form.

Column Title	Value 1	Value 2	Value 3
L_to (Euclidean distance from source area to target area)	0.30	0.35	0.44
L_centra (Distance of the source area to the city center, km)	0.24	0.24	0.24
L_bike_lane (Length of bike lanes, km)	0	0	0
...
t_0 (Travel time with free road network)	241.32	227.49	350.04
t_a (Travel time with load of the road network)	248.27	227.49	352.56
S_path (The length of the path along the road network)	0.56	0.45	0.60
k_1 Coefficient of uneven path	1.88	1.28	1.36
t_d (Travel delay time, s)	6.94	0.01	2.52
k_z (Congestion factor)	1.02	1.00	1.00

The dataset contains 139,342 rows (139,341 observations) and 57 columns corresponding to the number of dependent and independent variables.

Within the framework of the third dataset “Regions”, the influence of factors characterizing the parameters of the city area on the share of population trips from and to the area, as well as on the structure of these trips, was considered.

During the clustering of transport areas of the city, 24 factors from the dataset of the city macromodel were considered as independent variables, among which we can distinguish the distance of the source area from the city center, the number of exits and entries from/to the area, the number of workers in the area, the average travel time of different modes of transport, the average travel speed, etc.

As the desired value (dependent variable y), the share of trips by various types of transport from and to the area is determined, which is expressed as a percentage of the following values: the share of trips from the area by private transport ($from_car_%$); share of trips from the area by public transport ($from_bus_%$); the share of trips from the area by bicycle ($from_bike_%$); proportion of pedestrian trips from the area ($from_ped_%$); the share of trips to the area by private transport ($to_car_%$); share of trips to the area by public transport ($to_bus_%$); the share of trips to the area by bicycle ($to_bike_%$); share of pedestrian trips in the area ($to_ped_%$).

The dataset contains 400 rows (399 observations) and 32 columns corresponding to the number of dependent and independent variables.

2.2. Data Preparation

Technically, the process of developing neural network models was implemented using the Python programming language in the IPython programming environment using the Jupyter Notebook (V. 7.0.6). The initial data and their output were processed using the numpy and pandas frameworks. The results were visualized using the Matplotlib package and Seaborn. For the direct development of the model, the keras, tensorflow, and sklearn libraries were used. To process the data before creating the model, the StandardScaler module of the sklearn.preprocessing package was used. To split the data into training and test sets, the train_test_split module of the sklearn.model_selection package was used. For stepwise development of the model, the Sequential and Dense modules of the keras.models and keras.layers packages were used. The plot_model module of the tensorflow.keras.utils package was used to visualize the model architecture. To save and load neural network models and standardizers, the load_model modules of the keras.models package and the dump and load modules of the joblib package were used. To determine the accuracy metrics of the model, the r2_score, mean_squared_error, and explained_variance_score modules of the sklearn.metrics package were used.

Before further work with the data, the datasets were checked for empty values in the cells, and the rows with empty values were deleted using the df.isnull() and df.dropna() methods, respectively.

Before developing the architecture of the neural network and carrying out its training, the data were divided into the training and test sets, and the data were also standardized using the sklearn.preprocessing.StandardScaler module. Standardization involves scaling the data around zero, subtracting the sample mean from the actual value, and dividing the result by the standard deviation. Thus, all variables will have the same order of values before being placed in the model, which will improve the accuracy of the model.

The train_test_split function of the sklearn.model_selection module with test_size = 0.3, random_state = 42 parameters was used to form the training and test sets. For example, for the “Structure of mobility” dataset, the size of the training sample for independent variables (X) was 97,262 observations, and the test sample was 41,685 observations. The validation set was formed from the test set within the additional arguments of the model.fit function of the keras.models framework.

The procedure for working with standardized values is as follows: The actual values of the independent variables (X) of the training sample are standardized and then placed in the model for its training. The predictive value is also generated by the model in a standardized form and requires conversion to absolute values for its interpretation. To test the accuracy of the model, test values are also standardized and placed in the model to form a prediction. After the forecast for comparison with the initial data is received, the test values are standardized in the original form.

2.3. Development of a Neural Network

For all three tasks (three datasets), sequential-type models were implemented in the Python programming language using the keras library, as well as the sklearn library. Programmatically, the model was created by declaring the Sequential() model and sequentially adding layers through model.add using the Dense module from keras.layers. All three neural network models are aimed at solving multiple input factor regression problems with one or more dependent variables.

The task of the designed neural network for the dataset “Traffic” is predicting the values of the dependent variable characterizing the travel time in a loaded network from independent variables characterizing the influence of 32 values of traffic factors.

The neural network has 1 input layer with 32 neurons equal to the number of independent variables, 1 hidden layer with 40 neurons in it, and 1 output layer with 1 neuron equal to the number of independent variables. The model training parameters were chosen experimentally and are equal to the following values: loss = ‘mean_squared_error’, optimizer = ‘adam’, batch_size = 20, epochs = 50. The hidden layer activation function is relu.

The task of the designed neural network for the dataset “Mobility” is forecasting the values of dependent variables characterizing the structure of the mobility of the population of the area, in the amount of 4 items from 50 values of independent variables. The neural network has 1 input layer with 50 neurons equal to the number of independent variables, 1 hidden layer with 60 neurons in it, and 1 output layer with 4 neurons equal to the number of independent variables. The model training parameters were chosen experimentally and are equal to the following values: loss = ‘mean_squared_error’, optimizer = ‘adam’, batch_size = 20, epochs = 50. The hidden layer activation function is relu.

The task of the designed neural network for the dataset “Areas” is forecasting the values of dependent variables characterizing the share of trips of the population of the area by various modes of transport from the area and back, in the amount of 8 items from 24 values of independent variables. The neural network has 1 input layer with 24 neurons equal to the number of independent variables, 2 hidden layers with 40 neurons each and 1 output layer with 8 neurons equal to the number of independent variables. The training parameters of the model were chosen experimentally and are equal to the following values: loss = ‘mean_squared_error’, optimizer = ‘adam’, batch_size = 20, epochs = 50. The activation function for the 1st hidden layer is tanh, and that for the 2nd hidden layer is relu.

The architectures of neural networks developed for the conditions of this study are shown in Figure 2.

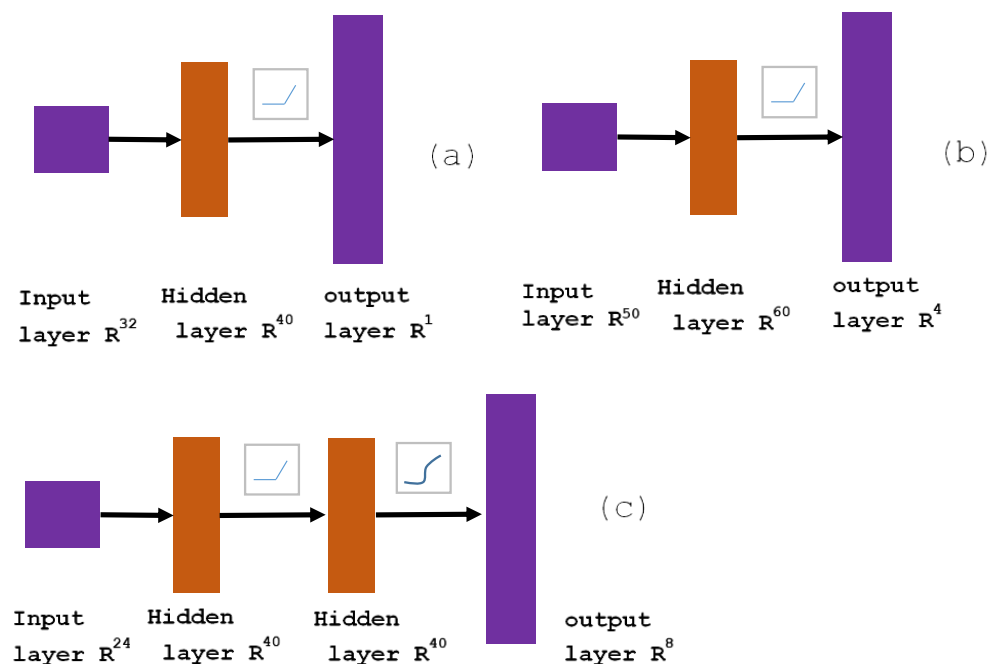


Figure 2. ANN architecture for determining (a) the influence of traffic factors on travel time in a busy network, (b) patterns of population mobility, and (c) the share of trips of the population of the area by various modes of transport from the area and back.

2.4. ANN Model Accuracy

The network model was trained on the dataset “Traffic” on the training sample using the above parameters and tested on the validation sample, as a result of which the model error significantly decreased, which indicates the correctness of the selected parameters and the completion of model training.

The model trained on the training set of data was used to predict the values of the mobility structure on the test set. The results of the first three values of the parameter to be predicted by the neural network—“travel time in a loaded network, s ”—are presented in Table 2, in which the second column is the actual test data t_a , and the third column is the data predicted by the model: t_a _pred.

Table 2. Comparison of test data (2nd column) and data predicted by the model (3rd column).

No.	ta	ta_pred
1	2776.8	2771.4
2	1442.8	1443
3	655	657
4	1393.7	1393.7
5	2366.1	2365.2

According to the correspondence of the values predicted by the model to the test ones, one can visually assess the accuracy of the model.

Model accuracy was quantified using the *r2_score*, *MSE*, and *explained_variance_score* metrics, which are commonly used to assess the accuracy of regression models. The results of determining the metrics are presented as follows: *r2_score*: 0.99; *MSE*: 2.65; *explained_variance_score*: 0.99. Based on the results of determining the quality metrics of the model, we can conclude that the mathematical model is statistically significant.

The model was trained on the Mobility dataset on the training set using the above parameters and tested on the validation set, as a result of which the model error significantly decreased, which indicates the correctness of the selected parameters and the completion of model training (Figure 3).

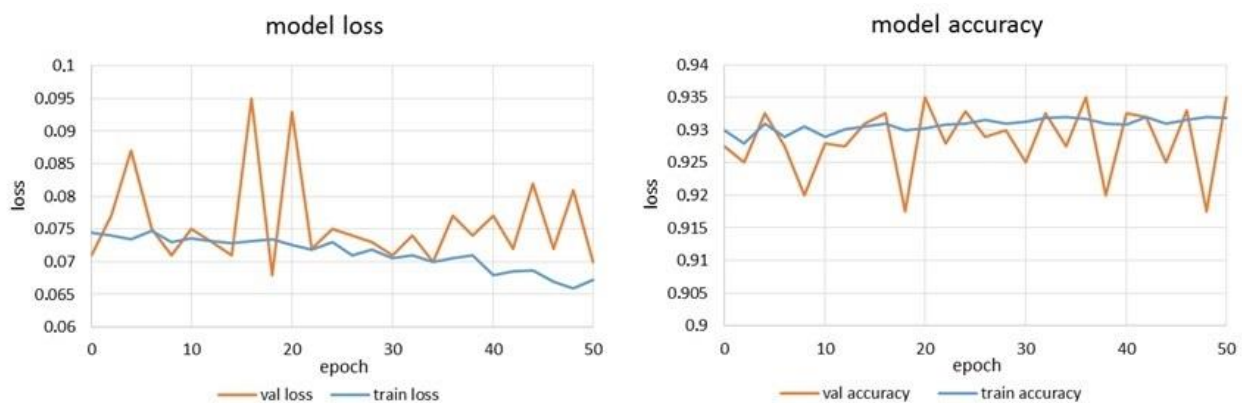


Figure 3. Change in the loss (errors) and accuracy (accuracy) metrics of the model for the training (blue) and validation (orange) samples according to the Mobility dataset.

The model trained on the training set of data was used to predict the values of the mobility structure on the test set. The results of the first three values predicted by the neural network are presented in Table 3, in which the first 4 columns are the actual test data, and the second 4 columns are the data predicted by the model: *car_pred*, *bus_pred*, *bike_pred*, *ped_pred*.

Table 3. Comparison of test data (columns 2–5) and data predicted by the model (columns 6–9).

No.	car_%	bus_%	bike_%	ped_%	car_pred	bus_pred	bike_pred	ped_pred
1	34.5	64.3	0.0	1.2	33.2	65.7	−0.4	1.5
2	27.8	48.8	2.2	21.3	29.8	50.7	2.3	17.7
3	91.7	4.9	0.0	3.4	91.2	6.2	0.0	3.0

All data in %.

According to the correspondence of the values predicted by the model to the test ones, one can visually assess the accuracy of the model.

Model accuracy was quantified using the *r2_score*, *MSE*, and *explained_variance_score* metrics, which are commonly used to assess the accuracy of regression models. The results of determining the metrics are presented as follows: *r2_score*: 0.92; *MSE*: 19.90; *explained_variance_score*: 0.92.

The model was trained on the dataset “Areas” on the training sample using the above parameters and tested on the validation sample, according to the results of which the model error significantly decreased, which indicates the correctness of the selected parameters and the completion of model training (Figure 4).

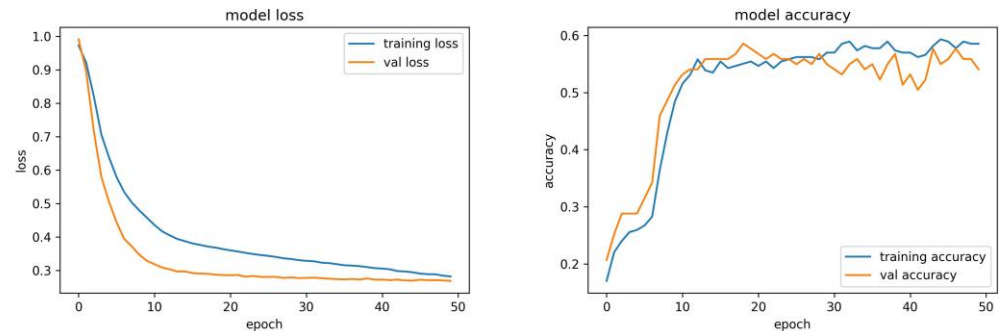


Figure 4. Change in the loss (errors) and accuracy (accuracy) metrics of the model for the training (blue) and validation (orange) samples for the “Areas” dataset.

After training and testing the artificial neural network model, a cluster analysis of city areas was performed. Cluster analysis was carried out to classify areas according to a number of criteria: distance from the city center, distance and time of travel by private and PT, parameters of transport demand. For cluster analysis, the self-organizing map method was used.

The self-organizing map method (SOM) or Kohonen self-organizing maps are a type of neural network algorithm. The SOM is an unsupervised learning method and, unlike neural networks trained by backpropagation, does not require labeled data to produce results. This class of methods is used, as a rule, for data classification or clustering problems when it is necessary to find a pattern in previously unknown data. The method allows you to transform data from a multidimensional space into a two-dimensional space that is convenient for perception and interpretation, creating a map or mask of the input data that allows you to understand their structure. The neural network within the framework of this method is a rectangular two-dimensional structure (map), the neurons of which are connected only to the input neurons, containing data samples from their dataset. A neuron is an n -dimensional column vector containing weights. In the process of training on input data, the weights of neurons (maps) are modified, trying to become as close as possible to the input vector according to the criterion of minimizing the square of the distance between the weight of the neuron and the value of the input data vector. The map neuron with the best result, which is closest to the input vector based on the totality of weights, is called the Best Matching Unit (winning neuron) or BMU. Thus, the entire map is “pulled” to the data point. For data clustering tasks, a self-organizing map was used with empirically determined parameters of vertical and horizontal dimensions equal to three, as well as with a dimension (number of features) of the input space equal to six.

To visualize the clustering of areas carried out using the method of self-organizing maps, the principal component method (PCA) was used, which makes it possible to visually represent multidimensional data in the space of two components that determine the dispersion of the input data, without reducing the quality of information. These components correlate with specific factors from the original dataset and can be conditionally interpreted as an integral indicator of the influence of these factors on the distribution of data by class. In the axes of the principal components, it is possible to see the most significant features of the source data, even despite the inevitable distortions.

3. Results

The model trained on the training set of data was used to predict the values of the mobility structure on the test set. The results of the first three values predicted by the neural

network are presented in Table 4, in which the first eight columns are the actual test data. Table 5 shows the second eight columns—the data predicted by the model: from_car (pred), from_bus (pred), from_bike (pred), from_ped (pred), to_car (pred), to_bus (pred), to_bike (pred), to_ped (pred).

Table 4. Test data from the sample to assess the accuracy of the model.

No.	fr_car	fr_bus	fr_bike	fr_ped	to_car	to_bus	to_bike	to_ped
1	51.4	44.4	0.2	3.9	46.9	49.5	0.3	3.3
2	29.4	38.1	0.5	31.9	32.7	46.7	0.7	20
3	26	30.1	0.4	43.5	33.9	45.8	0.3	19.9

All data in %.

Table 5. Data predicted by the model.

No.	fr_car (pred)	fr_bus (pred)	fr_bike (pred)	fr_ped (pred)	to_car (pred)	to_bus (pred)	to_bike (pred)	to_ped (pred)
1	58.3	37.2	0.8	4.7	56.1	39.3	0.3	3.7
2	28.1	37.4	0.5	34.0	32.1	45.2	0.7	21.8
3	25.7	29.1	0.5	44.8	32.9	41.5	1	23.8

All data in %.

The model accuracy was quantified using the *r2_score*, *MSE*, and *explained_variance_score* metrics, which are commonly used to assess the accuracy of regression models. The results of determining the metrics are presented as follows: *r2_score*: 0.68; *MSE*: 17.07; *explained_variance_score*: 0.69.

Based on the results of determining the quality metrics of the model, it can be concluded that the model is quite adequate and is able to predict values with an accuracy of about 70%. However, the model has the potential to improve the prediction accuracy by increasing the training dataset.

The results of the study are presented in several subsections and correspond to the three stages of the study, which are described in Section 2.

3.1. Impact of Travel Distance and Time on Mobility Structure

Using the data of the city transport model, dependences of the structure of transport mobility on travel time and length were obtained.

Graphs of the dependence of the share of trips by PT and private transport on the average travel time for each of 400 “target areas” of correspondence are shown in Figures 5 and 6.

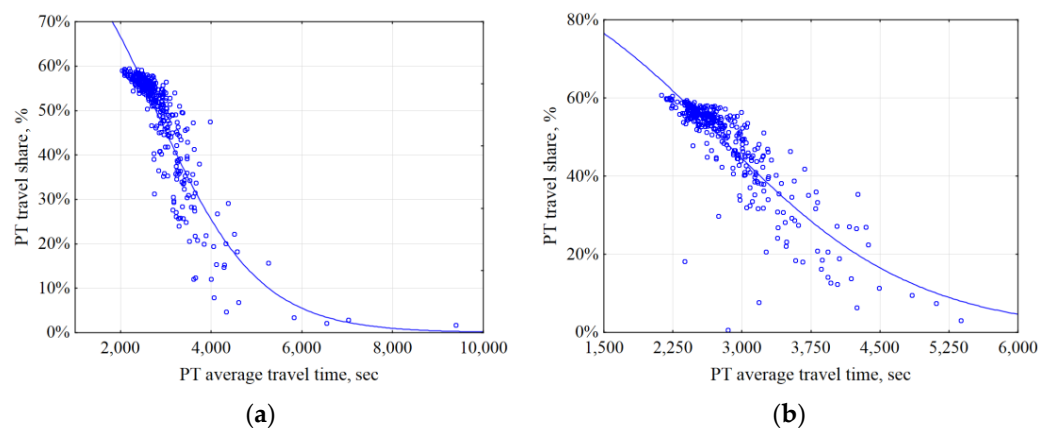


Figure 5. The influence of public transport travel time on the share of trips by public transport: (a) from the source areas; (b) to the target areas.

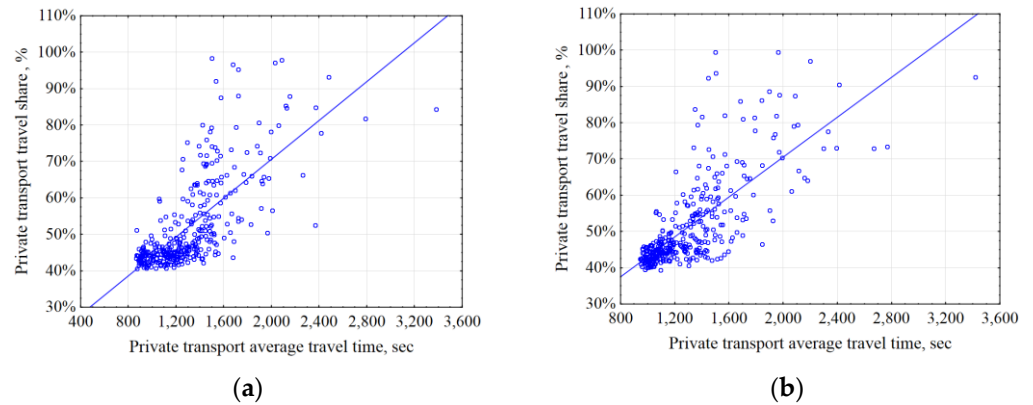


Figure 6. The influence of private transport travel time on the share of trips by private transport: (a) from the source areas; (b) to the target areas.

The results of the study confirm the hypotheses previously put forward in Section 2 about the type of mathematical models 1 and 2 of the influence of travel time on the share of trips by private and PT.

Figure 7 shows the shares of trips by private and PT depending on the travel distance along the road network.

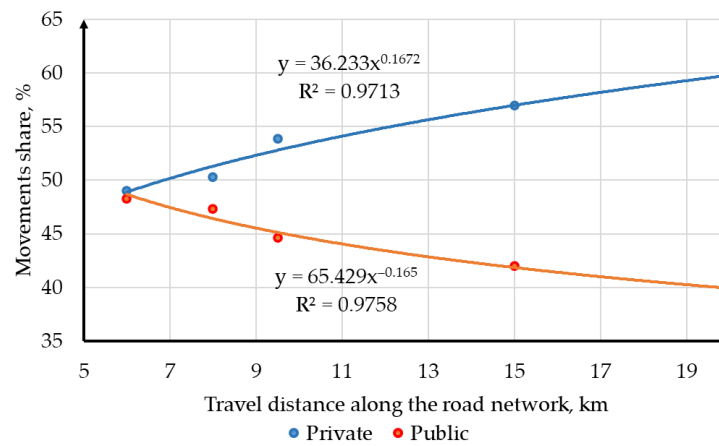


Figure 7. Shares of trips by private and PT depending on the travel distance along the road network.

This graph is based on average values for all 400 areas and correspondence, which leads to significant simplification. This dependence for the entire city does not allow us to identify problem areas, so the following are the results of the third stage of the study on clustering areas.

3.2. Clustering of City Transport Areas According to the Criterion of Population Mobility Structure

Data obtained for the population center as a whole provide incomplete information. For a more accurate assessment of population mobility in the city, source areas and target areas were divided into groups.

Using the k-means method, areas of the city transport macromodel were clustered. In this case, the accounting parameters used the shares of trips by cars and PT.

The clustering of transport areas in the macromodel of the city showed the division of areas into nine groups. Each group of districts differs in its characteristics (Tables 6 and 7) and distance from the city center. The largest number of areas in the central part of the city belonging to class 5 is due to their small area size.

Table 6. Transport area classification (source areas).

Area Group Class	Number of Areas per Class	Transport Area Parameters and Travel Characteristics (Source Areas)		
		Public Transport Share	Average Travel Distance by Private Transport, km	Average Travel Distance by PT, km
0	39	0.48	12.95	11.05
1	54	0.54	10.97	9.80
2	58	0.56	9.06	8.64
3	22	0.38	14.30	12.59
4	9	0.48	11.24	9.65
5	99	0.56	7.67	7.53
6	41	0.20	15.93	13.95
7	23	0.36	12.00	9.56
8	13	0.48	8.66	7.99

Table 7. Transport area classification (target areas).

Area Group Class	Number of Areas per Class	Transport Area Parameters and Travel Characteristics (Target Areas)		
		Public Transport Share	Average Travel Distance by Private Transport, km	Average Travel Distance by PT, km
0	38	0.265	17.09	14.76
1	23	0.382	14.22	10.95
2	26	0.423	13.06	9.99
3	37	0.455	12.49	11.49
4	28	0.512	11.3	9.58
5	20	0.487	9.66	8.57
6	31	0.516	10.11	9.54
7	54	0.55	9.06	8.54
8	100	0.564	7.68	7.45

All nine groups of transport areas can be divided into central, peripheral, and remote sectors. Examples of transport areas grouped into separate classes are shown in Figure 8.

Transport areas belonging to classes 0, 1, 4, and 7 (source areas) and 3, 4, 5, and 6 (target areas) are located on the periphery, outside the boundaries of the city central part, but within the bypass road (with rare exception).

Transport areas belonging to classes 2, 5, and 8 (source areas) and 7 and 8 (target areas) are located in the city central part with the largest population.

Transport areas belonging to classes 3 and 6 (source areas) and 0, 1, and 2 (target areas) are located either outside the bypass road or inside, but the bypass of the city is the border of the transport area.

After the transport areas were classified into two categories, namely source areas and target areas of correspondence, and the average values of output parameters and factors were determined for each class, graphs of the share of trips by mode of transport depending on the average travel time by private and PT were created (Figures 9–11).



Figure 8. Clustering of transport areas and division into sectors: (a) central sector (class 5, source); (b) central sector (class 8, target); (c) peripheral sector (class 1, source); (d) peripheral sector (class 7, target); (e) remote sector (class 6, source); (f) remote sector (class 0, target).

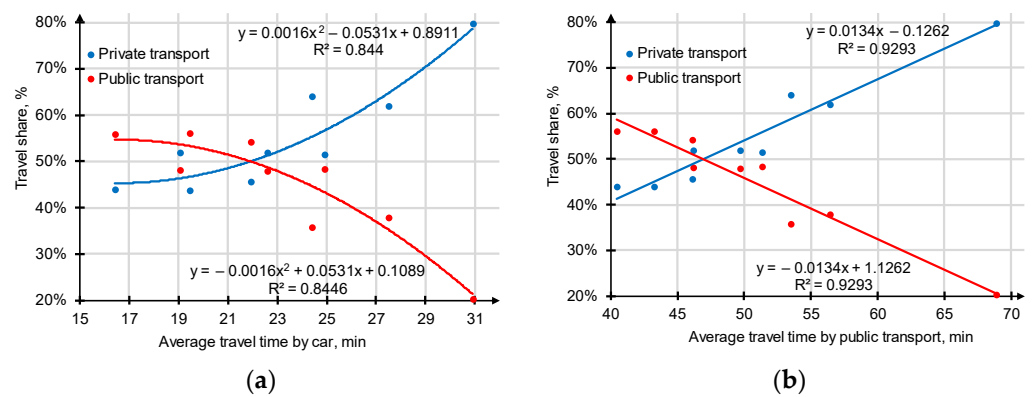


Figure 9. Average travel time influence on the share of travel by private and PT (source areas): (a) private transport; (b) PT.

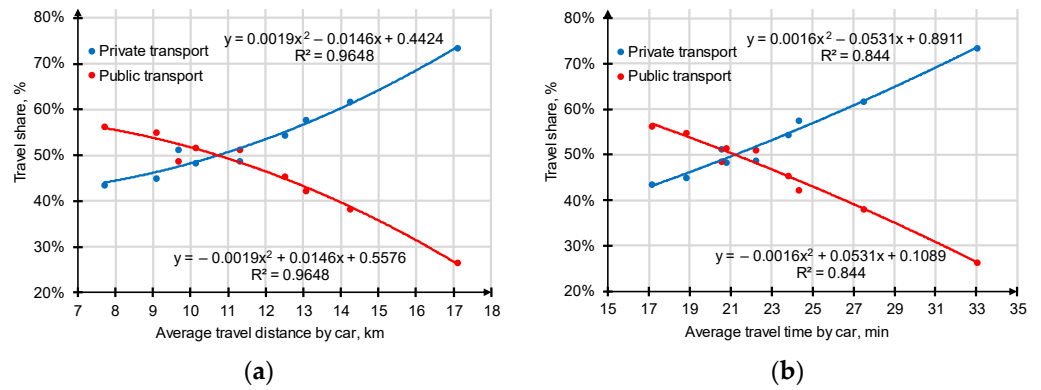


Figure 10. Average car travel time and travel distance influence on the share of travel by private and PT (target areas): (a) travel distance influence; (b) travel time influence.

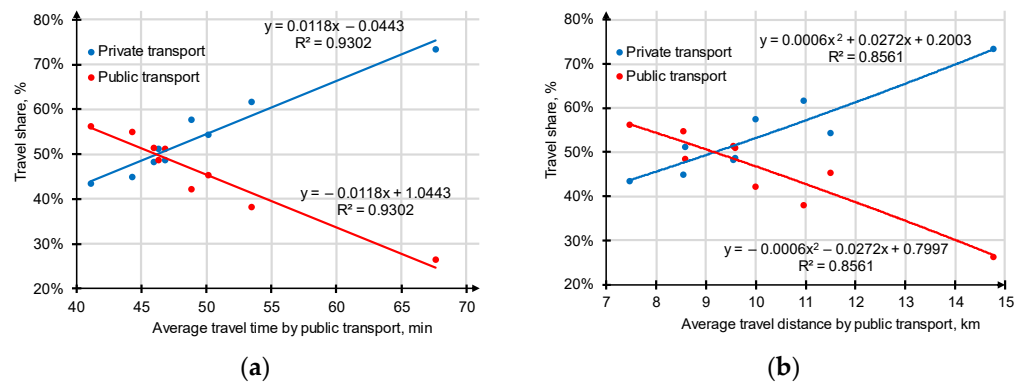


Figure 11. Average public transport travel time and travel distance influence on the share of travel by private and PT (target areas): (a) travel time influence; (b) travel distance influence.

As the transport area moves away from the city center and the travel time increases, the share of trips by car increases and by PT decreases. When using averaged parameters by groups of areas, the influence of the average travel time by private transport on the trips share is described by a quadratic model, and for PT, by a linear model. These research results can be used in the formation of a construction policy and the definition of a city development strategy—towards suburbanization and expansion of the city borders, with urbanization and an increase in the density of resettlement in the central part of the city or an increase in the density of resettlement in the peripheral areas of the city.

4. Discussion

A scientific problem has been solved to determine models and predict the influence of factors characterizing the transport infrastructure of a city area on the structure of the mobility of the population of this area and on traffic parameters, as well as on the share of population trips from and to the area and on the structure of these trips.

Neural network architectures have been developed to solve the regression problem for predicting the indicated parameters. The method of artificial neural networks with one (for predicting traffic parameters and mobility structure) or several hidden layers (for predicting the share of population trips from and to the area, as well as the patterns of these trips) was used. Neural networks are trained on a large sample of initial data obtained from the results of simulation macromodeling of urban transport network parameters and divided into three datasets: “Traffic”, “Mobility”, “Areas”.

Based on the results of determining the quality metrics of neural network models, their accuracy was determined, which is 99%, 92%, and 68%, respectively, for the datasets “Traffic”, “Mobility”, and “Areas”, from which we can conclude that they are adequate and have a high predictive value capability.

The above models of artificial neural networks were saved as separate files that store the architectures of the models and the weight coefficients of neurons. Pre-trained models can be used to predict the parameters of urban mobility without the need to form a macromodel of the city and conduct repeated simulations.

Based on the obtained artificial neural network, a program was developed to assess the structure of the mobility of the city population. The program is intended for the Administration of the city of Tyumen, subordinate industry institutions, and design organizations involved in the development and adjustment of transport planning documents, as well as the organization of transport services for the population of the city.

Transport areas belonging to the same class have different ratios of the share of trips by private and PT. Transport behavior is determined by a group of factors. Examples include the level of transport services for the population, taking into account the number of public transport routes, the coverage of the territory of a populated area by public transport routes in each area, and the number of multimodal routes that require a transfer to another mode of transport or to another route with re-payment of travel fares. For example, there are differences in the model of transport behavior in residential areas located at different distances from the city center. Thus, in the cottage village of Komarovo, located on the border with the bypass of the city of Tyumen, the share of trips by public transport is 0.55. For the village Derbyshi, which is located further from the bypass road, this figure is 0.4.

The ways of further research are determined and consist of the following:

- Clustering of correspondence on separate transport systems (private and PT) and determination of the most problematic PT routes and road network sections;
- Use of artificial neural networks in the transport planning of urban infrastructure, taking into account the economic factors of its organization.

The implementation of measures to ensure sustainable urban mobility requires financial resources from municipal or regional budgets. At the same time, the effectiveness of individual measures may differ significantly for transport areas, taking into account their location, and requires detailed elaboration in transport planning documents.

Author Contributions: Conceptualization, D.Z.; formal analysis, E.K. and A.P.; investigation, D.Z., A.P. and E.K.; project administration, D.Z. and D.C.; methodology, D.Z.; resources, D.C.; supervision, D.Z. and D.C.; validation, D.Z. and D.C.; data curation, D.C. and E.K.; funding acquisition, D.Z.; writing—review and editing, A.P. and D.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a grant from the Industrial University of Tyumen.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

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

References

1. Burrows, A.; Bradburn, J.; Cohen, T. *Journeys of the Future. Introducing Mobility as a Service*; Atkins Global: Epsom, UK, 2015; pp. 1–36.
2. Garrido-Jiménez, F.J.; Rodríguez-Rojas, M.I.; Vallecillos-Siles, M.R. Recovering Sustainable Mobility after COVID-19: The Case of Almeria (Spain). *Appl. Sci.* **2024**, *14*, 1258. [[CrossRef](#)]
3. Petrov, A.I.; Petrova, D.A. Sustainability of Transport System of Large Russian City in the Period of COVID-19: Methods and Results of Assessment. *Sustainability* **2020**, *12*, 7644. [[CrossRef](#)]
4. Domenteanu, A.; Delcea, C.; Chiriță, N.; Ioanăș, C. From Data to Insights: A Bibliometric Assessment of Agent-Based Modeling Applications in Transportation. *Appl. Sci.* **2023**, *13*, 12693. [[CrossRef](#)]
5. Arliansyah, J.; Prasetyo, M.R.; Kurnia, A.Y. Planning of City Transportation Infrastructure Based on Macro Simulation Model. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2017**, *7*, 1262–1267. [[CrossRef](#)]
6. Moscow Renovation Fund. Available online: <https://fr.mos.ru/> (accessed on 23 October 2023).
7. Zedgenizov, A.V.; I Solodkiy, A.; Efremenko, I. Assessment of suburbanized areas transport demand: Case study of the Irkutsk Agglomeration. *IOP Conf. Series Mater. Sci. Eng.* **2020**, *880*, 012075. [[CrossRef](#)]

8. Currie, G.; De Gruyter, C. Exploring links between the sustainability performance of urban public transport and land use in international cities. *J. Transp. Land Use* **2018**, *11*, 325–342. [[CrossRef](#)]
9. Almansoub, Y.; Zhong, M.; Raza, A.; Safdar, M.; Dahou, A.; Al-Qaness, M.A.A. Exploring the Effects of Transportation Supply on Mixed Land-Use at the Parcel Level. *Land* **2022**, *11*, 797. [[CrossRef](#)]
10. Bao, L.; Kusadokoro, M.; Chitose, A.; Chen, C. Development of socially sustainable transport research: A bibliometric and visualization analysis. *Travel Behav. Soc.* **2023**, *30*, 60–73. [[CrossRef](#)]
11. Yosifov, G.; Petrov, M. Review of Urban Traffic Detection Approaches with Accent of Transportation in Sofia, Bulgaria. Lecture Notes in Networks and Systems. In Proceedings of the 7th International Congress on Information and Communication Technology, London, UK, 21–24 February 2022; Volume 465, pp. 509–517.
12. Gammelli, D.; Yang, K.; Harrison, J.; Rodrigues, F.; Pereira, F.; Pavone, M. Graph Meta-Reinforcement Learning for Transferable Autonomous Mobility-on-Demand. In Proceedings of the 28th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2022, Washington, DC, USA, 14–18 August 2022; pp. 2913–2923. [[CrossRef](#)]
13. Ibáñez, L.-D.; Maddalena, E.; Gomer, R.; Simperl, E.; Zeni, M.; Bignotti, E.; Chenu-Abente, R.; Giunchiglia, F.; Westphal, P.; Stadler, C.; et al. QROWD—A Platform for Integrating Citizens in Smart City Data Analytics. *Stud. Comput. Intell.* **2023**, *942*, 285–321.
14. Weiwei, J. Bike sharing usage prediction with deep learning: A survey. *Neural Comput. Appl.* **2022**, *34*, 15369–15385.
15. Prado-Rujas, I.-I.; Serrano, E.; García-Dopico, A.; Córdoba, M.L.; Pérez, M.S. Combining heterogeneous data sources for spatio-temporal mobility demand forecasting. *Inf. Fusion* **2023**, *91*, 1–12. [[CrossRef](#)]
16. Nam, D.; Kim, H.; Cho, J.; Jayakrishnan, R. A model based on deep learning for predicting travel mode choice. In Proceedings of the Transportation Research Board 96th Annual Meeting Transportation Research Board, Washington, DC, USA, 8–12 January 2017.
17. Dabiri, S.; Heaslip, K. Inferring transportation modes from GPS trajectories using a convolutional neural network. *Transp. Res. Part C Emerg. Technol.* **2018**, *86*, 360–371. [[CrossRef](#)]
18. Zhang, J.; Zheng, Y.; Qi, D. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction. *Proc. AAAI Conf. Artif. Intell.* **2017**, *31*. [[CrossRef](#)]
19. Chung, J.; Sohn, K. Image-based learning to measure trac density using a deep convolutional neural network. *IEEE Trans. Intell. Transp. Syst.* **2017**, *19*, 1670–1675. [[CrossRef](#)]
20. Shepelev, V.; Aliukov, S.; Nikolskaya, K.; Shabiev, S. The Capacity of the Road Network: Data Collection and Statistical Analysis of Traffic Characteristics. *Energies* **2020**, *13*, 1765. [[CrossRef](#)]
21. Shepelev, V.; Aliukov, S.; Nikolskaya, K.; Das, A.; Slobodin, I. The Use of Multi-Sensor Video Surveillance System to Assess the Capacity of the Road Network. *Transp. Telecommun. J.* **2020**, *21*, 15–31. [[CrossRef](#)]
22. Zhang, L.; Zhang, G.; Liang, Z.; Ozioko, E.F. Multi-features taxi destination prediction with frequency domain processing. *PLoS ONE* **2018**, *13*, e0194629. [[CrossRef](#)] [[PubMed](#)]
23. Li, Z.; Yu, H.; Zhang, G.; Dong, S.; Xu, C.Z. Network-wide traffic signal control optimization using a multi-agent deep reinforcement learning. *Transp. Res. Part C Emerg. Technol.* **2021**, *125*, 103059. [[CrossRef](#)]
24. Yang, H.; Xie, K.; Ozbay, K.; Ma, Y.; Wang, Z. Use of Deep Learning to Predict Daily Usage of Bike Sharing Systems. *Transp. Res. Rec. J. Transp. Res. Board* **2018**, *2672*, 92–102. [[CrossRef](#)]
25. Debnath, A.; Singh, P.K.; Banerjee, S. Vehicular traffic noise modelling of urban area—A contouring and artificial neural network based approach. *Environ. Sci. Pollut. Res.* **2022**, *29*, 39948–39972. [[CrossRef](#)] [[PubMed](#)]
26. Li, C.; Bai, L.; Liu, W.; Yao, L.; Waller, S.T. Graph Neural Network for Robust Public Transit Demand Prediction. *IEEE Trans. Intell. Transp. Syst.* **2020**, *23*, 4086–4098. [[CrossRef](#)]
27. McMillan, L.; Varga, L. A review of the use of artificial intelligence methods in infrastructure systems. *Eng. Appl. Artif. Intell.* **2022**, *116*, 105472. [[CrossRef](#)]
28. Xiao, Z.; Fang, H.; Jiang, H.; Bai, J.; Havyarimana, V.; Chen, H. Understanding Urban Area Attractiveness Based on Private Car Trajectory Data Using a Deep Learning Approach. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 12343–12352. [[CrossRef](#)]
29. Liu, Q.; Yang, Z.; Cai, L. Traffic Flow Prediction of Expressway Section Based on RBF Neural Network Model Smart Innovation. Systems and Technologies. In Proceedings of the 5th International Conference on Wireless Communications and Applications 2021, Berlin, Germany, 17–19 December 2021; Volume 299, pp. 191–199.
30. Monje, L.; Carrasco, R.A.; Rosado, C.; Sánchez-Montañés, M. Deep Learning XAI for Bus Passenger Forecasting: A Use Case in Spain. *Mathematics* **2022**, *10*, 1428. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.