DOI: 10.23968/2500-0055-2025-10-1-89-102

ASSESSMENT OF THE TRANSPORT AND OPERATIONAL CONDITION OF ROADS BASED ON MOBILE LABORATORY DATA USING MACHINE LEARNING METHODS

Victoria Shamraeva

Financial University under the Government of the Russian Federation
49 Leningradsky Avenue, Moscow, 125993, Russia

*Corresponding author's email: VVShamraeva@fa.ru

Abstract
The subject of the study is the prediction of traffic intensity and pavement condition at a linear road section. The paper addresses a model of a neural network used to assess the usability of a road section and its transport and operational performance. The object of the study is a section of the M-1 Belarus road, 86th km, for the period from 2014 to 2023. The **purpose of the study** was to describe possible future scenarios of the road condition based on predicted traffic intensity and road quality condition metrics as part of the assessment of its usability with account for the International Roughness Index (IRI). In the course of the study, the following **methods** were used: Data Science (analysis of data collected from mobile laboratories) and machine learning algorithms (linear regression, gradient boosting, random forest, and neural networks based on long short-term memory (LSTM)). The output is a trained neural network capable of predicting the traffic intensity on the 86th km of the M-1 Belarus road. These methods reveal hidden patterns in the data and provide high-accuracy predictions. **Results**: The implementation of the deep learning model using the assessment of the condition of a linear road section will make it possible to address the main issues of road maintenance — to optimize time and reduce expenditures when planning and introducing measures at the stage of operation of transport infrastructure facilities, to take into account possible risks of road condition quality loss during re-pavement and design of new elements.

Keywords: transportation infrastructure; road diagnostics; defects; machine learning; Data Science; road information model.

Introduction

Road transport plays a paramount role in the transport infrastructure of the country. To ensure its development, the characteristics of roads must meet the traffic conditions since roads are subject to wear and tear as well as pavement deterioration due to heavy use and exposure to natural factors.

Nowadays, the role of digital technologies in business process management in almost all branches of technological industries has increased significantly. In this context, the road sector deserves special mention, where the growth rates of construction and reconstruction of linear road sections are quite high. For instance, the updated five-year road construction plan for 2024–2028 calls for 380 projects (see Resolution No. 3907-r dated 25.12.2023, http://government.ru/docs/50593/). These processes provide for an increase in the level of road maintenance, based on the diagnostics of linear road sections (Federal Law No. 257-FZ dated 08.11.2007 (amended on 04.08.2023) "On Motor Roads and Road Activities in the Russian Federation and Amendment of Certain Legislative Acts of the Russian Federation" (as amended and supplemented, effective from 01.03.2024), see https://legalacts.ru/doc/federalnyi-zakon-ot-08112007-n-257-fz-ob/, Article 14, Chapter 3).

Besides, switching to road construction financing with private investment (bonds and bank loans) has made it possible to increase the road network. Considering the lack of public resources, attraction of private investment requires the development of financial models that can guarantee payback at a required rate of return, which primarily depends on the correct prediction of traffic intensity on toll roads upon provision of comfort for road users. Thus, the compliance of pavement quality with the strict requirements for toll roads will increase the period between repairs and, therefore, optimize the repair budget, part of which can remain at the disposal of investors.

This paper is aimed to adapt machine learning methods to the prediction of the transport and operational condition of roads. For this purpose, approaches to pavement condition diagnostics are considered, the issues of predicting traffic intensity along a linear section and the influence of traffic on the transport and operational condition are addressed to determine the need for repair. Machine learning technologies open new horizons for road diagnostics goals and objectives. By applying these technologies, the expenses of companies can be significantly reduced due to timely diagnostics of road sections.

For citations: Shamraeva, V. (2025). Assessment of the transport and operational condition of roads based on mobile laboratory data using machine learning methods. Architecture and Engineering, No 1 (10), pp. 89–102. DOI: 10.23968/2500-0055-2025-10-1-89-102.

Methods

Both domestic (Apestin, 2011; Iliopolov et al., 2002; Vasilyev, 2013; and others) and foreign researchers (Panthi, 2009; Robinson et al., 1998; and others) have been studying the processes of road structure condition degradation. In the listed works, these processes are based on two levels (project level and network level (Fig. 1)) and two main indicators of pavement condition (operational and structural indicators (Fig. 2)). The project level is limited to the pavement design stage, while the network level considers pavement conditions throughout the entire life of the pavement. All stages of the life cycle of a road facility in the context of BIM modeling of linear road sections are described in (Shamraeva and Savinov, 2021).

The Russian Highways State Company adheres to a different methodology when assessing the actual level of the transport and operational condition of roads, which is based on the assessment of the residual life (Shamraeva, 2020) of road structures and division of operated road sections into three levels (standard-compliant, satisfactory and unsatisfactory) of pavement preservation (Fig. 3).

According to the regulatory requirements of Industrial Road Standards ODN 218.6.039–2018 "Recommendations on diagnostics and assessment of the technical condition of motor roads" (see https://files.stroyinf.ru/Data2/1/4294845/4294845825.pdf), the main transport and operational indicators of a road are as follows: speed provided by a road, traffic capacity, capacity for traffic of vehicles and road trains with axial and total weight established for the relevant categories of roads, etc. (Pugachev et al., 2024). All these indicators are fully reflected in such integrated indicator as traffic speed, expressed through the coefficient of design speed provision. As it is known, the transport and operational indicators of linear road sections are ensured by the parameters of longitudinal and transverse profiles, pavement strength, pavement evenness and adhesion characteristics, etc. (Figs. 1–3). Therefore, according to the methodology of ODM 218.4.039–2018 (Rosavtodor, 2018), the preservation of a road section depends on the average visual assessment score for pavement condition and the actual strength factor — the International Roughness Index (IRI) of non-rigid pavement.

Intensive road operation increases the risk of traffic accidents, reduces traffic safety, results in longer travel time and higher costs of transport repair and maintenance, etc. In this regard, it is important to

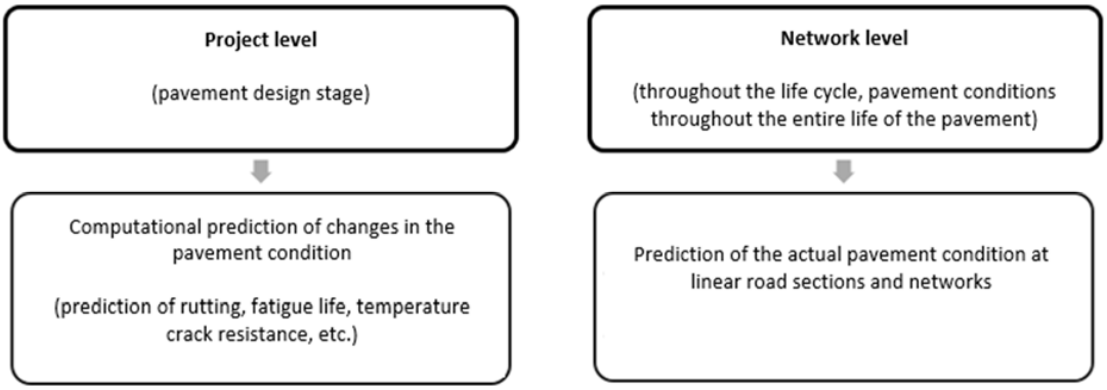

Fig. 1. Levels of road structure condition degradation processes

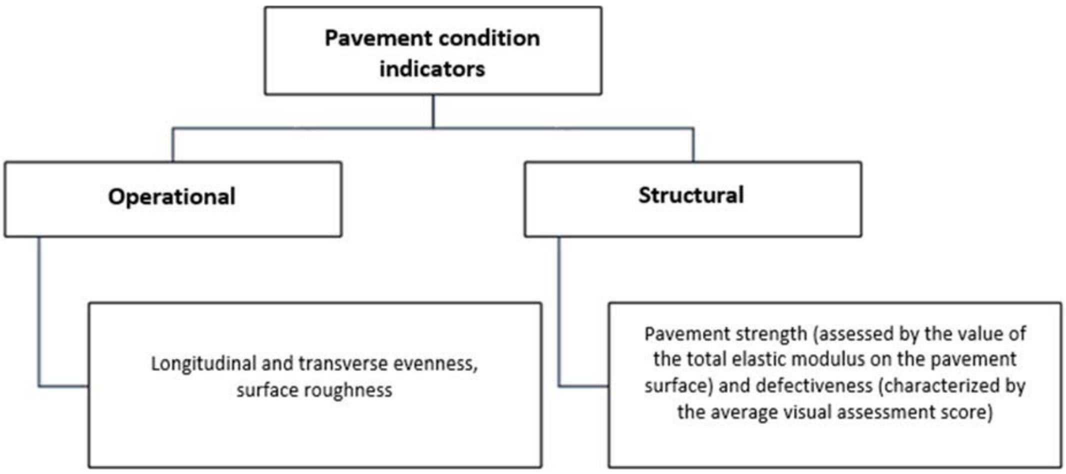

Fig. 2. Main groups of pavement condition indicators

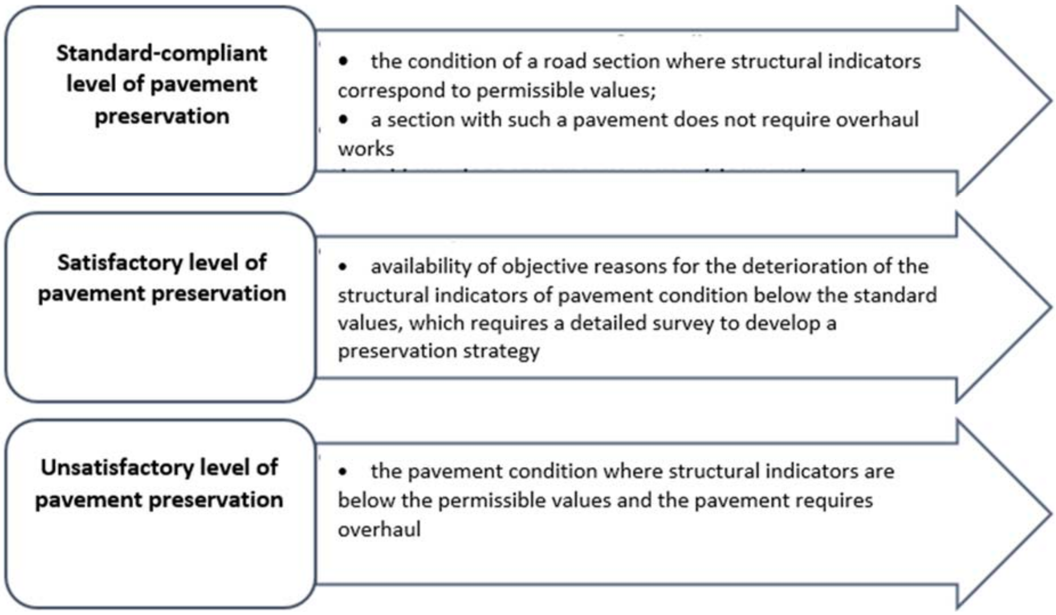

Fig. 3. Levels of pavement preservation

have in place effective methods for the diagnostics and assessment of the technical condition of roads. Transport and operational condition is assessed by specialized engineering organizations using mobile diagnostic laboratories, certified devices and equipment. Mobile units are equipped with modern devices and equipment (Znobishchev and Shamraeva, 2019) and may include panoramic video cameras of the latest generation (Stolbov et al., 2017) and even artificial intelligence technologies. Machine learning methods can be an alternative tool for monitoring the condition of road facilities (Marcelino et al., 2019).

The object of the study is a section of the M-1 Belarus road, 86th km, and its transport and operational condition. The aim was to develop a model for the diagnostics and assessment of the technical condition of roads using machine learning algorithms and medium-term prediction in accordance with the requirements of Organization Standard AVTODOR 2.28-2016 “Predicting the condition of operated roads of the Avtodor State Company” (Order No. 67 dated 06.05.2016) (Avtodor, 2016).

Input data processing

Two datasets are used as input data:

1 — data on traffic intensity with division into six categories of vehicles in the period from 2014 to 2023 inclusive in tabular format (Microsoft Excel), obtained from automated traffic counting points of the Avtodor State Company. The data are presented for each category of transport (cars, vans, light-duty trucks, single-unit vehicles, buses, road trains up to 13 m, from 13 to 18 m, and over 18 m) in forward and backward directions, including the total for two directions. In addition to the total intensity at a section for a certain period, the table uses the following structure: first the intensity is plotted by hours within one day, then the information for a day is summarized, and thus the data for a year are combined within one table sheet. As a result, information on traffic intensity for the past 10 years of observations for one section of the M-1 road is obtained.

2 — values of longitudinal evenness IRI by year for each road section from the 84th to 95th km of the M-1 Belarus road, obtained during measurements performed by a mobile diagnostic laboratory.

Following the basic idea of machine learning, it is necessary to provide a large number of proper examples based on which the program will “learn”. Therefore, it is important to check the input data for completeness so that they include all possible cases. Errors are critical since they will cause the model to give incorrect answers. The balance of the original dataset is also essential. If some cases are predominant, the model will favor the majority when making predictions. Therefore, when enough data are collected, it is necessary to pre-process them. The logic of the pre-processing of the intensity data from the first dataset is as follows: 1) the file is loaded and the names of all the table sheets in the sheet_ names (‘2014–2023’) list are read; 2) an empty combined_df dataframe is initialized to store the combined data from all the sheets, while the dataset is cleared of unnecessary data attributes. At the beginning of the cleanup function running, all sheets of the dataset table are visited, and the list of intensity data objects for the year of type pd.DataFrame is finally formed from the Microsoft Excel data source (Pandas library method — pd.read_excel()) into the combined_df variable, as well as unnecessary upper informative rows are deleted and a column with the date of the traffic intensity observation within the accuracy of one day is added; 3) all elements of the list are combined into a large single dataframe with the logical deletion of the index column where the observation date was stored. This combined_df dataframe is the result of the pre-processing of the input dataset with data on traffic intensity for the past 10 years of expert observations; 4) the processed dataframe is saved to a new Excel file.

The intensity trend can vary from year to year for various global reasons: social, economic, or political. For example, the COVID virus affected the life of the country for several years and reduced the mobility of its inhabitants, thus affecting the intensity indicator. Month is also an important attribute that affects intensity. In some months, commercial traffic is increasing: for instance, many people travel by car to neighboring countries during summer. Intensity will also vary depending on the day of the week. For example, on weekdays/weekends, people may have different reasons for using the M-1 road. Thus, the model requires attributes from the date and values of total traffic intensity in forward and backward directions. Before moving on to this task, it is also important to understand the amount of the output data and whether there is a way to effectively reduce it in terms of dimensions without loss of quality. One of such ways described in this paper is to group all traffic intensity indicators (regardless of vehicle type) by date column, which is already represented and ready to be grouped within the accuracy of one day using the Pandas library capabilities, indicating the summation of intensity indicators for all hours within a day, which is very convenient in computation and informative in presentation. The final step in the dataset pre-processing is the selection of new temporal attributes of the dataset, necessary for further training of the neural network. Particularly in this case, the following attributes are selected: month number, day number, day number of the week, and year of traffic intensity observation. When predicting the road evenness, the number of vehicles that have traveled along the road over a year is used. Therefore, the date and intensity for that date will be used to develop the model. For each road that goes forward and backward, corresponding values shall be saved. Then the total count will be excessive and should not be kept.

The correlation matrix presented in Fig. 4 shows the relationships between different attributes, including forward and backward movement as well as time parameters (year, month, day of the week, day of the month). There is a correlation of 0.18 between forward and backward movement, indicating a weak positive relationship. This means that some tendency for an increase in forward movement is accompanied by an increase in backward movement. The correlation between year/month and traffic intensity is not significant (0.18 for year and month with forward movement; –0.14 and 0.14 for backward movement). This may indicate that there are no clear trends by year or month in the data, or that any trends are non-linear and are not captured by linear correlation. The correlations with

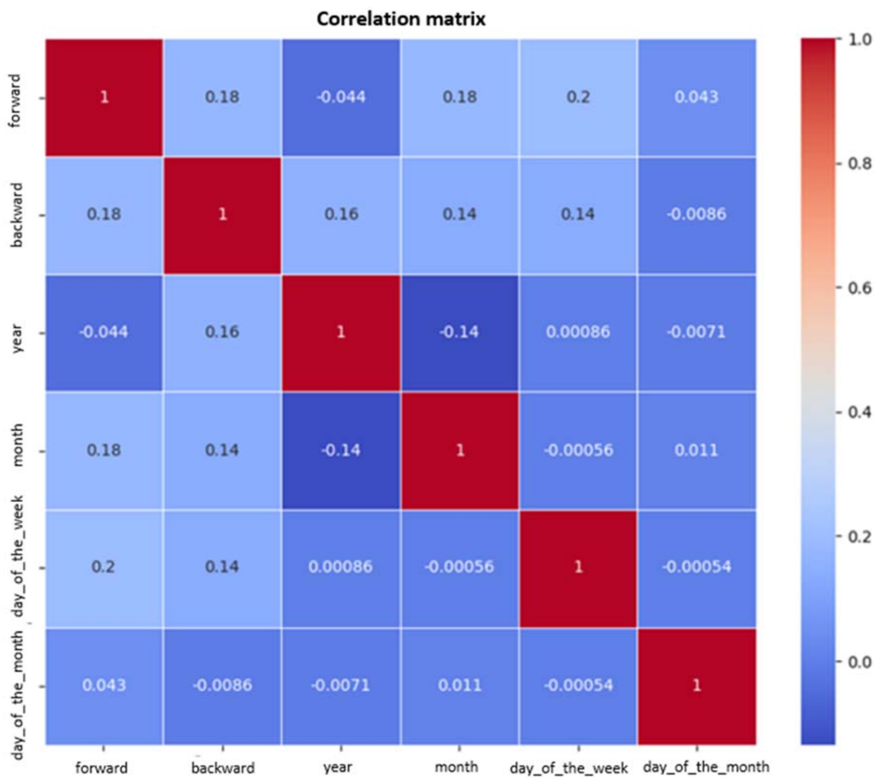

Fig. 4. Matrix of correlations between forward and backward traffic intensity and time

such attributes as the day of the week and the day of the month are also extremely low, indicating that there is no direct relationship of the day of the week or the day of the month with traffic intensity. When forming the output for a machine learning model, it is important to take into account that temporal attributes can significantly non-linearly affect the results. Despite the weak correlation between the forward and backward directions, these attributes can be useful when considered in conjunction with other variables that may strengthen their relationship or provide additional context.

Using machine learning methods that can capture such dependencies may yield more useful results. Such models as ensemble decision trees (e.g., random forest, gradient boosting) and neural networks (Goodfellow et al., 2018) can better handle such data as they are capable of detecting complex patterns and interactions between variables (Lasisi and Attoh-Okine, 2018).

In the second dataset with the IRI evenness values, a violation of the requirements of GOST 33388–2015 (Automobile roads of the general use. Requirements to conducting diagnostics and certification) (Interstate Council for Standardization, Metrology and Certification (ISC), 2016) can be observed, where the IRI value should not exceed 3.1 (https://nevacert.ru/dokumenty/normadoc/standartsbase-gost/gost-33388-2015). For instance, for 2023, such data are shaded in red in Fig. 5. In this file, the column with the GOST 33388-2015 standard values is redundant, not suitable for model training, and can be deleted during data pre-processing. Visualization of changes

	A	B	C	D	E	F	G
1	2023						
2	Beginning of the section, km	End of the section, km	Forward direction		Backward direction		GOST 33388-2015 standard value
3			Lane 1	Lane 2	Lane 3	Lane 4	
4	84,000	84,100	2,7	2,5	1,7	2,0	3,1
5	84,100	84,200	2,4	1,9	1,6	2,3	3,1
6	84,200	84,300	2,0	1,8	1,8	1,8	3,1
7	84,300	84,400	2,4	2,0	1,6	1,8	3,1
8	84,400	84,500	1,7	1,8	1,4	1,9	3,1
9	84,500	84,600	1,9	1,7	1,2	1,8	3,1
10	84,600	84,700	2,1	1,9	1,4	1,5	3,1
11	84,700	84,800	1,9	1,6	1,2	1,5	3,1
12	84,800	84,900	1,9	1,9	1,1	1,4	3,1
13	84,900	85,000	2,0	1,6	1,2	2,0	3,1
14	85,000	85,100	2,5	1,7	1,2	1,3	3,1
15	85,100	85,200	2,4	1,8	1,1	1,2	3,1
16	85,200	85,300	2,0	2,2	1,1	1,4	3,1
17	85,300	85,400	2,8	3,3	1,0	1,7	3,1
18	85,400	85,500	3,9	2,6	1,1	1,7	3,1
19	85,500	85,600	2,4	2,2	1,3	2,6	3,1
20	85,600	85,700	2,8	3,5	1,3	2,2	3,1
21	85,700	85,800	3,0	3,1	2,4	2,3	3,1
22	85,800	85,900	2,0	1,7	2,3	2,1	3,1
23	85,900	86,000	2,7	1,8	2,1	2,3	3,1
24	86,000	86,100	2,6	1,8	2,0	1,8	3,1
25	86,100	86,200	2,1	1,9	2,0	2,2	3,1
26	86,200	86,300	3,4	2,1	1,8	2,5	3,1
27	86,300	86,400	3,6	2,6	2,4	2,3	3,1

Sheet1 Sheet2

Fig. 5. IRI values at different sections of M-1 for 2023

in the mean IRI values for each lane (Fig. 6) shows that the road condition deteriorated over time.

For example, for lane 3 (Fig. 6), a slight asymmetry to the right is observed, where the bulk of the data is between 1 and 4. It can also be seen that the values did not change in the two years from 2019 to 2020

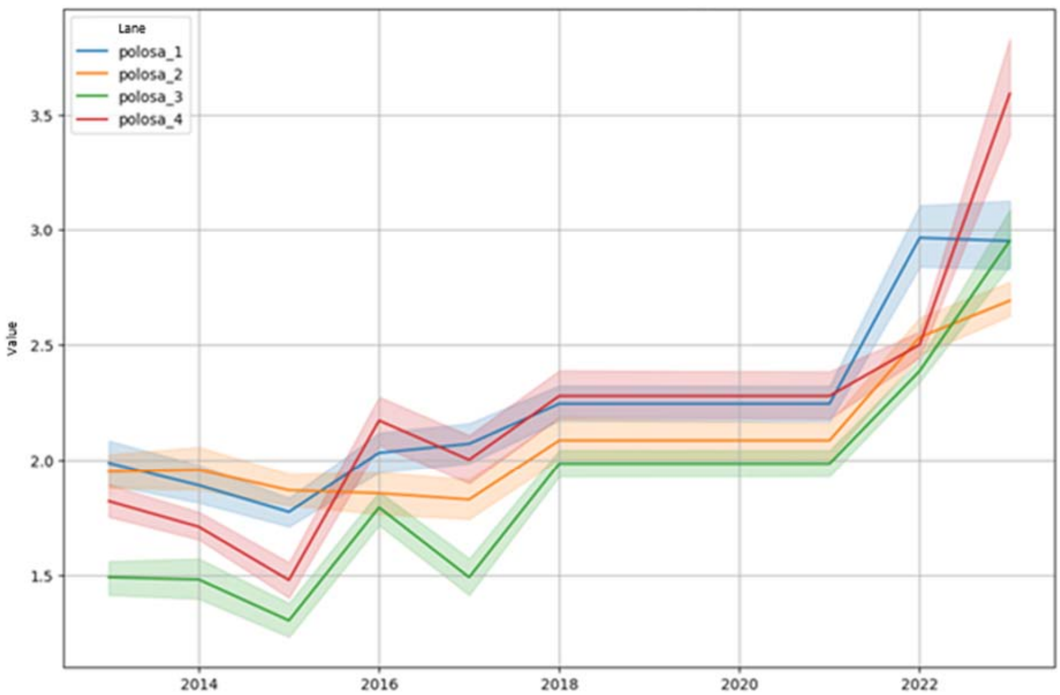

Fig. 6. Diagram of changes in the mean IRI values for each lane

(data for 2019–2020 were not available in the dataset provided). To determine the law of traffic intensity distribution for each lane, the SciPy library and the stats module of the dynamic Python programming language were used, which provides functions and statistical methods for work with probability distributions and statistical tests. For example, for lane 3 (Fig. 7), comparison of the values of the statistics of various tests (Kolmogorov–Smirnov, Pearson, Shapiro–Wilk, etc.) with the smallest p-value of the significance test equal to 0.05, resulted in a conclusion that the observed data are consistent with the Gumbel distribution on the right. As for other lanes, based on checks by statistical tests, it was concluded that the annual distributions of the IRI values are suitable for approximation modeling with subsequent transfer to machine learning models. The final IRI dataframe is shown in Fig. 8.

Training model selection and tuning

The preparatory stages of model building and training, with their further evaluation for the purposes of predicting values based on the second dataset with pre-processed data are as follows:

1. Break down the data sample into attributes and target values.
2. Select the attributes and variables based on which the target values will be predicted.
3. Divide the sample into training and test samples (it should be noted that for correct operation, the test sample should not be used for training).

This will result in an optimal prediction model for different target columns based on historical data and its application for further prediction for the period 2024–2034.

Let us briefly highlight the main points of the process of creating a suitable prediction model for different target columns.

Predicting changes in the IRI indicator

Creating model training functions: Here the data are split into test and training samples. Dataframe Y contains the values based on which the model will make predictions. These are the “start” values (Fig. 8) for the beginning of the road segment. Since all segments are of the same length, this value is sufficient. Year values are also used to track changes in pavement quality over the course of a year. Dataframe Y contains the IRI values themselves. Then, according to the principle of machine learning, the data are split into test and training samples in the proportion of 4 to 1. This is the optimal value since it allows for splitting so that there are more data left for training, but there are also data left for validation.

Model selection, training, tuning and evaluation of results: Model selection is carried out depending on the task (Cano-Ortiz et al., 2022), e.g.: classification, regression, clustering. Examples of models are decision tree, neural networks, linear regressions, support vector methods, etc. (Hosmer et al., 2013). Once the model is defined, it is trained, tuned, and the results are evaluated. If necessary, if the results are unsatisfactory, the model is reconfigured. Besides, when model training is too long or imprecise, it is possible to simplify or complicate it. When the resulting model shows good results, it can be used for practical tasks, and then it is implemented into the original system for which it was created (Gazarov, 2020).

A loss function is used to evaluate the model. With its help, the model “understands” how far it is from the ideal state. The difference between the answers or the resulting values is calculated. Once the loss function has been determined, it is necessary to understand how to fix the model, how to change the weights properly to improve further results. Optimization algorithms are used for this purpose. The main method to minimize the loss function is gradient descent, which, by gradually changing the parameters in the direction, reduces the error. XGBoost is a powerful machine learning model based on the gradient boosting method. It operates by starting from creating a simple model, such as a decision tree, for initial predictions. Next, prediction errors are obtained and new trees trained on the error

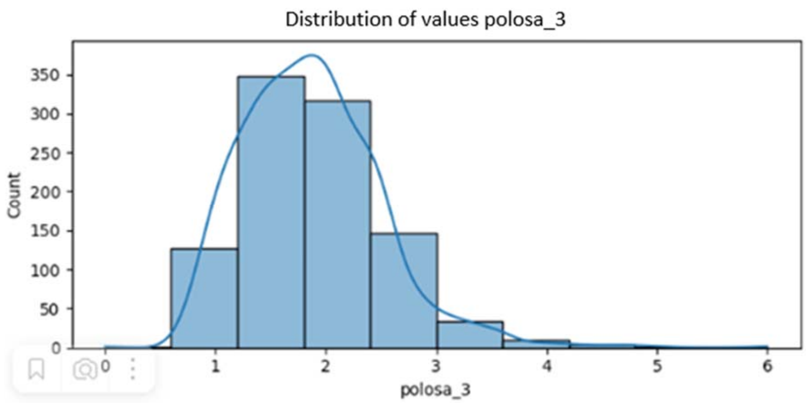

Fig. 7. Distribution of values polosa_3

	start	end	polosa_1	polosa_2	polosa_3	polosa_4	year
0	84.0	84.1	2.7	2.5	1.7	2.0	2013
1	84.1	84.2	2.4	1.9	1.6	2.3	2013
2	84.2	84.3	2.0	1.8	1.8	1.8	2013
3	84.3	84.4	2.4	2.0	1.6	1.8	2013
4	84.4	84.5	1.7	1.8	1.4	1.9	2013
...	...	...	...	...	...	...	...
985	94.5	94.6	2.5	2.3	3.2	2.8	2023
986	94.6	94.7	3.3	2.5	5.0	2.9	2023
987	94.7	94.8	2.8	2.4	4.7	3.1	2023
988	94.8	94.9	2.7	2.4	3.0	2.9	2023
989	94.9	95.0	2.7	2.5	2.9	2.5	2023

Fig. 8. IRI dataframe

are added based on them. These iterations continue until a given number of cycles have been executed or the error is small. When regulation and other optimization methods are used, the model shows good results. Random forest is an ensemble-based machine learning model that differs from the previous one in that it operates by creating multiple decision trees and combining their predictions. Each tree is trained on a random subsample of the input data, which promotes model diversity. Besides, each node partitioning step in the tree uses a random subset of attributes, which adds additional randomness and reduces the correlation between the trees. Once all trees are trained, the model predicts the result for a new observation by aggregating the predictions of all trees: in the case of the classification task, voting is applied, where the final class is determined by a majority vote, and in the regression task — the predictions are averaged. This algorithm is more robust and can show good results too.

The above models are trained on training data and their performance is evaluated on a test sample using the mean_squared_error and r2_score metrics (Shamraeva, 2024). The results are stored in the "results" dictionary. The attribute matrix X includes start and year columns, and the target variables are represented in data columns polosa_1, polosa_2, polosa_3 and polosa_4 (Fig. 8). For each target variable, models are trained and evaluated separately. The results are then displayed on the screen.

Creating a prediction model: New future_years data were created to predict values for the period 2024–2034. New future_years years were generated using the "range" function, and the structure of the new Dataframe was created so that each year is repeated for all unique values of the "start" column from the original data (Fig. 8). This preserves the consistency between the "start" values and the new years. Predictions were made using trained models. The results of the predictions are stored in the "predictions" dictionary.

To further visualize the predictions, graphs were constructed where the years are plotted on the X axis and the predicted values for each target variable are plotted on the Y axis. The mean prediction values for each year were calculated using the "groupby" method and plotted in the graphs. Thus, this process involves data loading, model training and evaluation, and prediction of future values followed by visualization of the results.

The above models showed unsatisfactory results and are not suitable for this task. Another method — linear regression — was tried.

Linear regression is a machine learning method that models linear relationships between a dependent (target) variable and one or more independent variables (attributes) (Hosmer et al., 2013). The model tries to find the best straight line that minimizes the sum of squares of errors among the predicted and actual values of the target variable. This can be achieved by estimating the coefficients of a linear equation, where each coefficient indicates how much the target variable will change if the corresponding attribute is changed by one unit. The training process involves finding coefficients that minimize the MSE using optimization techniques such as the least squares method. As a result, the MSE value amounts to 0.2667. This means that the mean square deviation of the predicted values from the actual ones is 0.2667. The values of the target variable vary within a small range, such as 2 to 6, therefore, the error of 0.2667 may be significant. This shows that this model did quite well in terms of results. However, this is not the best way to predict the IRI evenness — the traffic intensity prediction according to Organization Standard AVTODOR 2.28-2016 should be taken into account.

Predicting changes in traffic intensity

In this study, simple machine learning models, such as linear regression, random forest, gradient boosting, were used for each road direction. After training, visualization of the prediction for each movement was made (Fig. 9). Fig. 9 shows the results of the predictions of different traffic intensity models for the period from 2024 to 2034 covering forward movement. By examining each graph in Fig. 9, we can draw the following conclusions:

1. Linear regression (the model is represented by the blue line) shows the lowest accuracy of predictions; large deviations at the beginning and end of the period are especially noticeable.

2. Random forest (the model is represented by the orange line) shows more accurate predictions compared to linear regression but still has some deviations from the actual values, especially at peak times.

3. Gradient boosting (the model is represented by the green line) most accurately reproduces seasonal variations in traffic intensity and deviates least from the actual values.

It can be seen from the graph (Fig. 9) that traffic intensity has a pronounced seasonal nature with peaks and declines that are most accurately predicted by the gradient boosting model. A neural network was used to refine the results. For its setting, the values are balanced so that no attribute would dominate by its increased mean value.

In the course of this study, a long short-term memory (LSTM) model was created and trained. The LSTM model is a recurrent neural network (RNN) designed to efficiently train and predict time series and sequential data. It is capable of memorizing important information in long sequences and forgetting unnecessary information due to its special architecture consisting of memory cells controlling input, output and forget gates. The input gates decide which information from the current input to keep, the forget gates determine which information to delete from the memory cell, and the output gates regulate which information from the cell will be used to compute the output. This structure allows the LSTM to capture and preserve long-term dependencies. For the prediction task, the LSTM model includes two LSTM layers. The first LSTM layer contains 50 neurons and returns sequences, the second LSTM layer also contains 50 neurons. The LSTM layers are followed by a dense layer with 50 neurons and an activation function (rectifier linear unit (ReLU)), followed by a dense output layer with two neurons to predict two target variables. The results of the neural network training showed that the model was trained for 100 epochs, where 72 steps were performed at each epoch at a rate of about 3 milliseconds per step. In the last few epochs, the loss values in the training sample ranged between 0.0094 and 0.0098, while the loss values in the validation sample remained stable, fluctuating between 0.0106 and 0.0121. The final loss in the test sample was 0.0091, indicating that the model performed well with the training, validation and test samples with minimal variation in loss values.

Next, prediction was made for ten years. The graph (Fig. 10) shows complex relationships with peaks in summer months and declines in winter months. Traffic intensity in both forward and backward directions maintains an overall upward trend. This indicates a projected increase in traffic flow in the future.

Analysis of the results

Prediction of pavement technical condition assessment was made using two approaches:

- based on past years' IRI measurements using a linear regression model,

- based on equation 6.1 of Organization Standard AVTODOR 2.28-2016, using the traffic intensity of future years that was predicted by a neural network model.

Using Python code, predictions of the IRI values for the given years (Fig. 11a) and traffic intensity values (Fig. 11b) were generated using a machine learning model. The results of intensity value predictions are displayed on the screen, showing the date, intensity in forward and backward direction (Fig. 11b).

In this form, the data are ready to plot graphs of monthly forward and backward traffic (Fig. 12) and

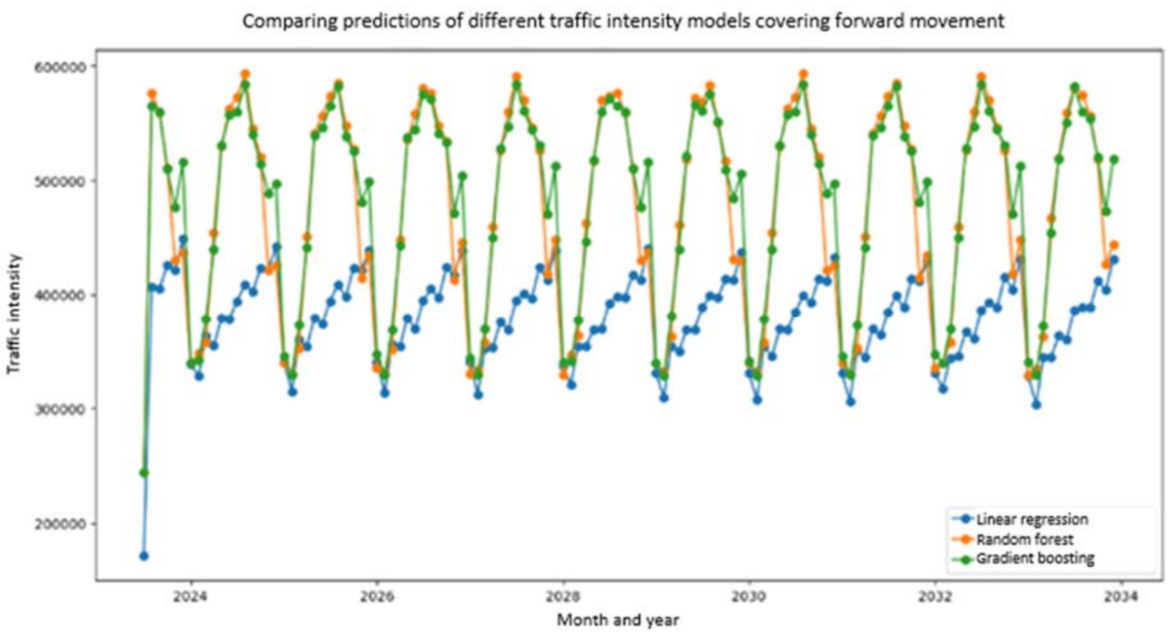

Fig. 9. Comparing predictions of different traffic intensity models covering forward movement

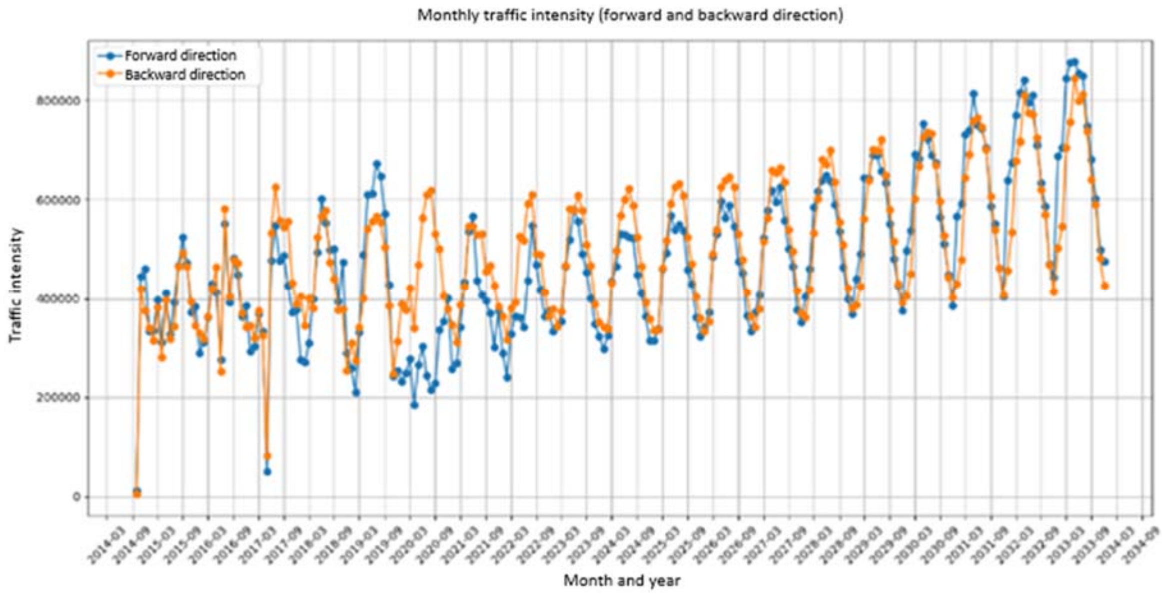

Fig. 10. Intensity prediction

	start	year	polosa_1	polosa_2	polosa_3	polosa_4
0	84.0	2024	3.170604	2.876721	2.659333	3.280827
1	84.1	2024	3.166205	2.871084	2.660291	3.278382
2	84.2	2024	3.161805	2.865447	2.661250	3.275937
3	84.3	2024	3.157406	2.859810	2.662209	3.273491
4	84.4	2024	3.153006	2.854174	2.663168	3.271046
...	...	...	...	...	...	...
1095	94.5	2033	3.688758	2.950728	3.955234	4.344944
1096	94.6	2033	3.684359	2.945091	3.956193	4.342499
1097	94.7	2033	3.679959	2.939454	3.957152	4.340054
1098	94.8	2033	3.675560	2.933817	3.958111	4.337608
1099	94.9	2033	3.671160	2.928181	3.959070	4.335163

a) Prediction results: IRI

131/131 [==============================] - 0s 1ms/step

	дата	прямое	обратное
0	2023-07-19	13300.680664	16415.083984
1	2023-07-20	16052.410156	15689.279297
2	2023-07-21	21617.058594	15172.226562
3	2023-07-22	19045.304688	17619.160156
4	2023-07-23	8680.833008	25879.925781
...	...	...	...
4179	2034-12-27	17137.482422	12002.160156
4180	2034-12-28	20424.916016	12052.699219
4181	2034-12-29	17821.062500	13173.458984
4182	2034-12-30	12690.513672	16863.751953
4183	2034-12-31	2375.298828	26900.529297

b) Predicting traffic intensity

Fig. 11. Prediction results

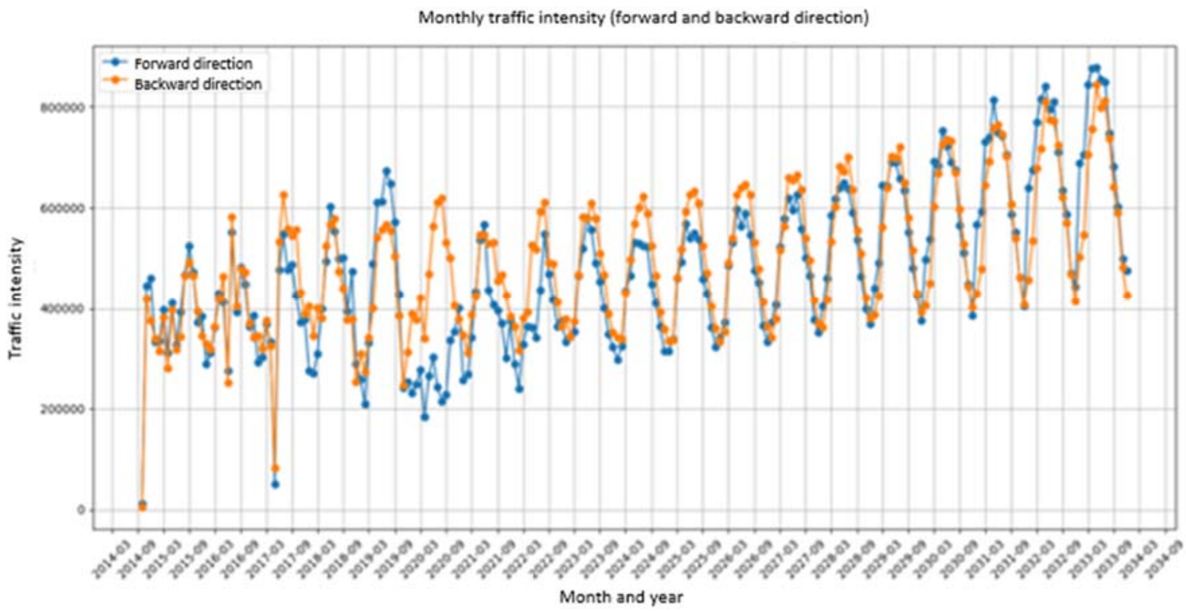

Fig. 12. Predictions of monthly traffic intensity in forward and backward directions through March 2034

draw conclusions with subsequent findings to make recommendations.

The prediction section (Fig. 12) shows that seasonal fluctuations continue, with peaks in summer months and declines in winter months. Traffic intensity in both forward and backward directions maintains an overall upward trend. This indicates a projected increase in traffic flow in the future. The forward (blue line) and backward movement (orange line) continue to show parallel trends with similar seasonal amplitudes. Despite year-to-year fluctuations, the overall trend is upward, which could indicate an expected increase in population or economic activity, or improved infrastructure. In 2033 and 2034, traffic intensity reaches values of approx. 800,000, which is significantly higher than the values in previous years. In general, the forecast indicates a steady increase in transportation intensity despite annual seasonal fluctuations, suggesting the need for further development of transport infrastructure.

To ensure a high level of M-1 Belarus road maintenance, it is necessary to implement measures for the development of transport infrastructure: road reconstruction, including increasing the number of lanes and constructing new interchanges (which will help distribute traffic flows and reduce the load on the existing infrastructure sections), preventive pavement repair to maintain the standard IRI value and ensure high average traffic speed. The analysis also revealed that the composition of traffic would not change. This route is characterized by an equal proportion of various means of transport. Fig. 13 shows a 10-year forecast for the IRI indicator using machine learning for four lanes of the road. Each of the four bar graphs shows the average IRI coefficient for the corresponding lane for the period from 2024 through 2033. Fig. 13a shows that IRI for the first lane gradually increases from 2024 to 2033, starting at about level 3 and reaching a value just above 4. This indicates a gradual deterioration of the pavement condition on the first lane. Fig. 13b shows a similar trend for the second lane, where IRI also gradually increases from just below 3 in 2024 to around 3.5 in 2033, also indicating an increase in pavement roughness. In Fig. 13c, the forecast for the third lane shows an increase in the IRI values from 2024 to 2033, starting at about 2.5 and reaching a value of about 3.5, indicating a deterioration of pavement quality on this lane. Fig. 13d shows an increase in IRI for the fourth lane, starting at just above 2.5 in 2024 and reaching critical values of 4 in 2033, which also indicates significant pavement deterioration.

Fig. 14 shows the 10-year IRI coefficient prediction for the four road lanes on the 86 km of the M-1 highway. Each line on the graph corresponds to one of the lanes: the blue line represents the first lane, the orange — the second, the green — the third, and the red — the fourth. The graph covers the period from 2024 to 2034, indicating a gradual deterioration of the pavement condition.

On the first lane (blue line), IRI increases from approx. 3.0 to 4.5. On the second lane (orange line), IRI starts around 2.8 and reaches approx. 4.2. The third lane (green line) shows an increase in IRI from 3.2 to 4.7. On the fourth lane (red line), IRI values start around 3.5 and increase to 5.5 by 2034. The red dashed line at level 4 on the Y axis indicates the maximum permissible IRI value. All four lanes show an uptrend. Fig. 14 shows that the IRI coefficients for all lanes tend to increase over time. Lane 4 (red line) already exceeded the maximum permissible value and continues to grow faster than the others. Lanes 1, 2, and 3 (blue, orange, and green lines, respectively) also show growth but currently remain below acceptable levels. All lanes show a linear increase in the IRI coefficient, indicating continuous pavement deterioration over the next ten years. The highest rate of deterioration is observed in the fourth lane, while the lowest start and end points are characteristic of the second lane. This highlights the importance of planning financial expenditures for pavement improvement and rehabilitation to slow down the development of bumps and preserve road quality.

In addition to assessing the IRI longitudinal evenness, it is necessary to take into account transverse evenness (rutting) as well since this defect affects the steering ability and assignment of repairs (due to pavement wear). Using equation

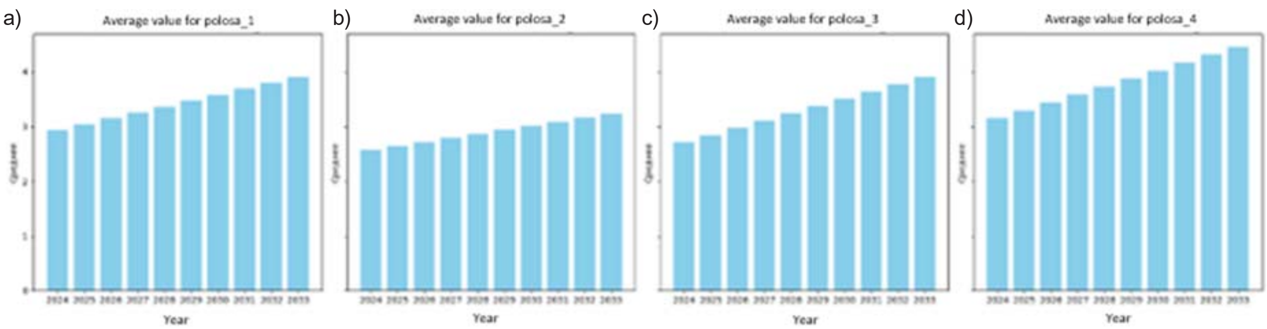

Fig. 13. Predicting the IRI coefficient for 10 years by machine learning

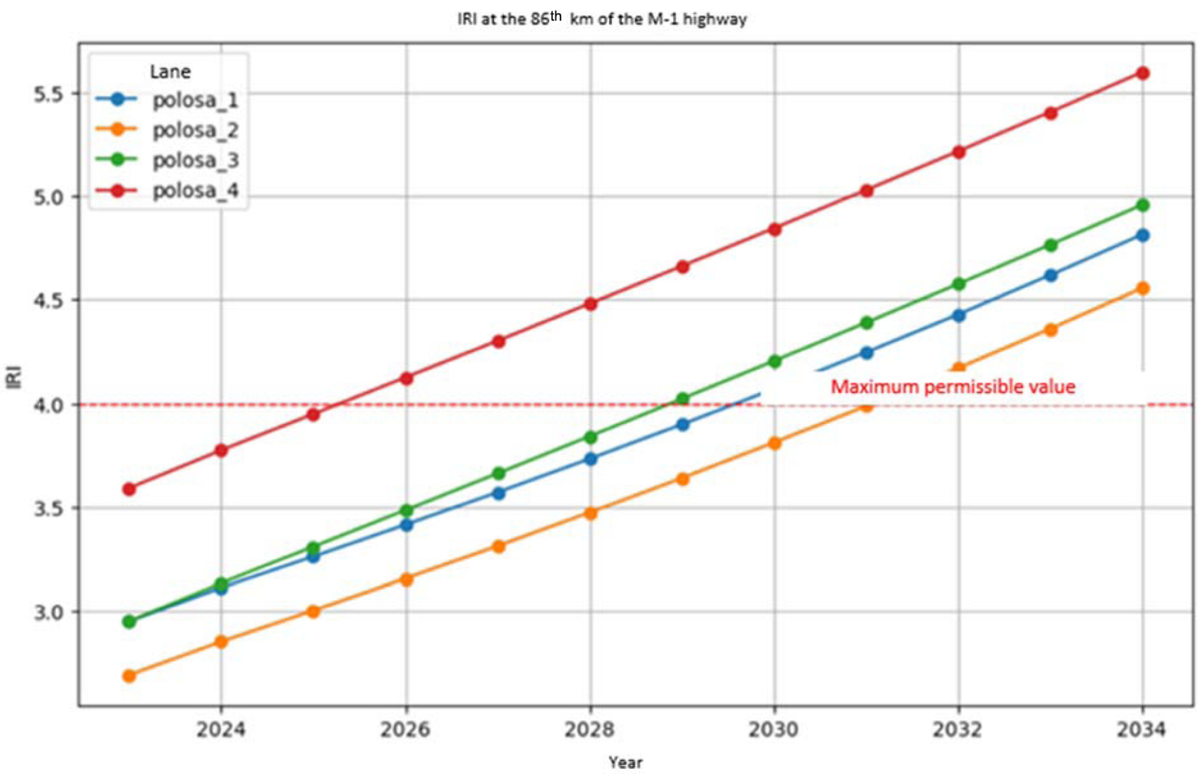

Fig. 14. Traffic intensity prediction: 10-year changes in IRI

6.1 of Organization Standard AVTODOR 2.28-2016 for rutting calculation, changes in transverse evenness (rutting) were forecast (Fig. 15). The graph in Fig. 15 represents a forecast for rutting on the 86 km of the M-1 highway for 10 years into the future. It shows four different lanes, each marked with a different color: blue, orange, green, and red, respectively. The horizontal axis indicates the years from 2024 to 2034, and the vertical axis shows the rutting coefficient in cm. The lines on the graph

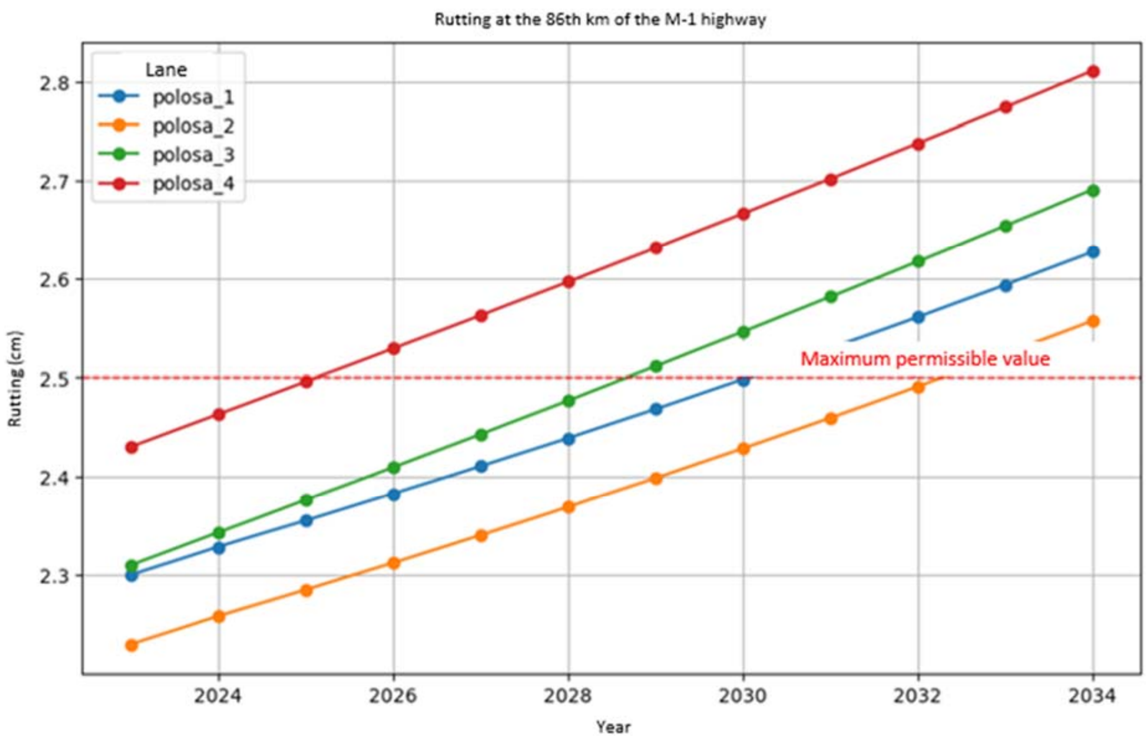

Fig. 15. Forecast of rutting for 10 years

show a linear trend of increasing rutting, indicating continuous pavement deterioration over the forecast period. The fourth lane has the highest level of rutting and shows the greatest rate of increase, which may indicate heavy load or wear. All lanes show a steady increase in rutting over time, indicating a gradual deterioration of the pavement condition.

The first lane (blue line) starts at about 2.4 cm in 2024 and reaches about 2.6 cm by 2032. The second lane (orange line) shows growth from 2.45 cm to just above 2.65 cm over the same period. The third lane (green line) starts around 2.5 cm and grows to about 2.75 cm. The fourth lane (red line) has the highest initial level, about 2.6 cm, and reaches a level of about 2.85 cm by 2032. The red dashed line indicates the maximum permissible rutting value of 2.5 cm. The polosa_4 lane, indicated by the red line, is already exceeding this value by approx. 2025. Lanes 3 and 1, indicated by the green and blue lines, respectively, reach the limit value near the end of the forecast period. Lane 2, indicated by the orange line, remains below the maximum permissible value throughout the forecast period, but also shows a significant increase.

All of the above graphs (Figs. 14–15) emphasize the need for regular pavement monitoring as well as maintenance and repair to slow down the deterioration of the road and ensure safe and comfortable traffic on the M-1 highway. The fourth lane could be out of service in 2025. The first and third lanes expect maintenance in 2029. The second lane will remain intact for ten years.

Conclusion

The relevance of the work is determined by the need to develop an effective strategy for the operation of the road network, ensuring the continuity and safety of traffic. The main objective of the study was to develop an optimal road network repair strategy considering road conditions based on machine learning methods. As a result of this study, the road owner will receive detailed and optimal cost schedules for the planned periods of pavement operation and repair for several years in advance. In addition, the customer will be given the opportunity to make a more accurate calculation of the optimal budget for each stage of repair, allowing for effective financial planning. It can be applied by road authorities to develop a variety of effective solutions that can be easily adapted to new conditions and parameters and will contribute to the rational use of resources and minimize subjective factors in the maintenance planning process. The study has theoretical significance in the development of a road infrastructure condition management strategy model using machine learning methods. Thanks to more accurate and timely diagnostics of the technical condition of the road, it is possible to carry out rehabilitation and repair in advance, which in turn will reduce the cost of overhauls and extend the life of the road.

The implemented deep learning model can be improved, modernized and updated. For instance, one of the development directions could be the use of other neural network architectures or the use of a different type (recurrent or convolutional networks) (Cha et al., 2017) in order to apply other approaches and mathematical algorithms in value prediction tasks (Zari et al., 2015). Besides, this model has the possibility to be extended by adding new class and functional components based on the new assigned tasks and functions in the field of road transportation. These results and basic reasoning show the maximum utility of modern machine learning technologies on real-life examples in the subject area, which once again indicates the importance of developing this direction in the IT sphere in general.

References

Apestin, V. K. (2011). About divergence between design and normative interrepair periods of road pavement service. *Science and Engineering for Highways*, No. 1, pp. 18–20.

Avtodor (2016). *Avtodor State Company Standard. Predicting the condition of operated roads of the Avtodor State Company. Organization Standard AVTODOR 2.28-2016*. Moscow: Avtodor, 26 p.

Cano-Ortiz, S., Pascual-Muñoz, P., and Castro-Fresno, D. (2022). Machine learning algorithms for monitoring pavement performance. *Automation in Construction*. Vol. 139, 104309. DOI: 10.1016/j.autcon.2022.104309.

Cha, Y.-J., Choi, W., and Büyüköztürk, O. (2017). Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 32, Issue 5, pp. 361–378. DOI: 10.1111/mice.12263.

Gazarov, A. R. (2020). Advantages of using artificial intelligence in the field of construction. *News of the Tula State University. Technical Sciences*, No. 4, pp. 136–139.

Goodfellow, I., Bengio, Y., and Courville, A. (2018). *Deep learning*. 2nd edition. Moscow: DMK Press, 652 p.

Hosmer Jr., D. W., Lemeshov, S., and Sturdivant, R. X. (2013). *Applied logistic regression*. 3rd edition. New York: John Wiley & Sons, 528 p.

Iliopolov, S. K., Seleznev, M. G., and Uglova, Ye. V. (2002). *Dynamics of road structures*. Rostov-on-Don: Rostov State University of Civil Engineering, 258 p.

Interstate Council for Standardization, Metrology and Certification (ISC) (2016). *GOST 33388-2015. Automobile roads of the general use. Requirements to conducting diagnostics and certification*. Moscow: Standartinform, 11 p.

Lasisi, A. and Attoh-Okine, N. (2018). Principal components analysis and track quality index: a machine learning approach. *Transportation Research Part C: Emerging Technologies*, Vol. 91, pp. 230–248. DOI: 10.1016/j.trc.2018.04.001.

Marcelino, P., de Lurdes Antunes, M., Fortunato, E., and Castilho Gomes, M. (2019). Machine learning approach for pavement performance prediction. *International Journal of Pavement Engineering*, Vol. 22, Issue 3, pp. 341–354. DOI: 10.1080/10298436.2019.1609673.

Panthi, K. (2009). *A methodological framework for modeling pavement maintenance costs for projects with performance-based contracts*. PhD Thesis in Civil Engineering. DOI: 10.25148/etd.FI09120824.

Pugachev, I. N., Sheshera, N. G., Grigorov, D. E., and Evtyukov, S. S. (2024). Forecast of traffic flow intensity. Training with a teacher. Random tree method. *T-Comm*, Vol. 18, No. 4, pp. 36–47. DOI: 10.36724/2072-8735-2024-18-4-36-47.

Robinson, R., Danielson, U., and Snaith, M. (1998). *Road maintenance management: concepts and systems*. London: Red Globe Press, 312 p. DOI: 10.1007/978-1-349-14676-5.

Rosavtodor (2018). *Industry road guidance document. Recommendations on diagnostics and assessment of the technical condition of motor roads. ODM 218.4.039-2018*. Moscow: Rosavtodor, 59 p.

Shamraeva, V. V. (2020). Economic efficiency of operation of elements of civil buildings taking into account residual resource: probabilistic-statistical approach. *Modern Science: Actual Problems of Theory and Practice. Series: Economics and Law*, No. 2, pp. 61–68.

Shamraeva, V. V. (2024). Mathematical methods for forecasting stock price changes and their implementation using machine learning methods. *Fundamental Research*, No. 11, pp. 88–96. DOI: 10.17513/fr.43718.

Shamraeva, V. and Savinov, E. (2021). INFRA-BIM for business processes' management in road construction and operation. *Architecture and Engineering*, Vol. 6, No. 3, pp. 19–28. DOI: 10.23968/2500-0055-2021-6-3-19-28.

Stolbov, Yu. V., Stolbova, S. Yu., Pronina, L. A., and Uvarov, A. I. (2017). The provision of the inspections' control accuracy on the IV, V category road basics and coverings by the application of H-3 type niveliers. *Russian Automobile and Highway Industry Journal*, No. 6 (58), pp. 125–132.

Vasilyev, A.P. (2013). *Operation of motor roads*. In 2 volumes. Vol. 2. 3rd edition. Moscow: Academia Publishing Center, 320 p.

Ziari, H., Sobhani, J., Ayoubinejad, J., and Hartmann, T. (2015). Prediction of IRI in short and long terms for flexible pavements: ANN and GMDH methods. *International Journal of Pavement Engineering*, Vol. 17, Issue 9, pp. 776–788. DOI: 10.1080/10298436.2015.1019498.

Znobishchev, S. and Shamraeva, V. (2019). Practical use of BIM modeling for road infrastructure facilities. *Architecture and Engineering*, Vol. 4, No. 3, pp. 49–54. DOI: 10.23968/2500-0055-2019-4-3-49-54.

ОЦЕНКА ТРАНСПОРТНО-ЭКСПЛУАТАЦИОННОГО СОСТОЯНИЯ ДОРОГ НА ОСНОВЕ ДАННЫХ ПЕРЕДВИЖНЫХ ЛАБОРАТОРИЙ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

Виктория Шамраева

Финансовый университет при Правительстве Российской Федерации,
125993, Ленинградский пр-т., 49, Москва, Россия

E-mail: VVShamraeva@fa.ru

Аннотация
Введение. Предметом исследования является прогноз интенсивности движения транспортных средств и состояние дорожного покрытия на линейно протяженном участке автомобильной дороги. В статье рассмотрена модель нейронной сети, посредством которой произведена оценка потребительских свойств участка дороги и его транспортно-эксплуатационных показателей. Объектом исследования является участок дороги М-1 Беларусь, 86-й километр за временной промежуток с 2014 по 2023 годы. **Цель исследования:** описание возможных будущих сценариев состояния дороги, исходя из предсказанных показателей транспортной интенсивности и метрик состояния качества дороги в рамках оценки её потребительских свойств с учётом ровности по показателю International Roughness Index (IRI). **Методы**: Data Science (анализ данных), собранных с передвижных лабораторий, а также алгоритмы машинного обучения (линейная регрессия, градиентный бустинг, случайный лес и нейронные сети на основе Long Short-Term Memory (LSTM)). На выходе создана обученная нейронная сеть, способная спрогнозировать интенсивность транспортного движения на 86 км дороги М-1 Беларусь. Эти методы позволяют выявить скрытые закономерности в данных и обеспечить высокую точность прогнозов. **Результаты:** реализация модели глубокого обучения на примере оценки состояния линейного участка автомобильной дороги позволит решать основные задачи содержания дороги - оптимизировать время и средства при планировании и реализации мероприятий на этапе эксплуатации объектов транспортной инфраструктуры, учитывать возможные риски потери качества состояния дороги при её обновлении и проектировании новых элементов.

Ключевые слова: транспортная инфраструктура; диагностика автомобильных дорог; дефекты; машинное обучение; Data Science; информационная модель дороги.