Journal of Urban and Environmental Engineering, v.19, n.1, p.17-32
ISSN 1982-3932
doi: 10.4090/jueev19n1.2025.017032

Journal of Urban and Environmental Engineering
www.journal-uee.org

DEEP LEARNING NEURAL NETWORK MODEL FOR PREDICTING RISKS AND TIMING OF ROAD INCIDENTS

Anton Finogeev*, Maxim Polezhaev, Alexander Grushevsky and Alexey Finogeev
Penza State University, 40 Krasnaya St., Penza, 440026, Russian Federation

Received 24 May 2025; received in revised form 16 June 2025; accepted 18 June 2025

Abstract: One of the ways to improve road safety is a predictive analysis of the risks of critical events depending on the dynamics of changes in accident rates, traffic violations, and the influence of objective and subjective factors. The analysis of accident rates and factors influencing the risks of road incidents is performed by studying the data presented in the form of time series using a deep learning neural network model. A type of recurrent neural network model with LSTM layers, a transformer, and a multi-headed attention mechanism is proposed as a tool for predictive analysis of time series. Time series are formed based on data extracted from photo radar systems for photo and video recording in the road transport environment. In the course of the study, a methodology for training a model for predicting the activation time of critical incidents in the road transport environment was developed through the analysis of time series of accident rates, traffic violations, and a number of objective influencing factors. The article presents the results of training and research of model variants for choosing its structure and selecting parameters in order to obtain an acceptable forecast error. The article presents studies of the influence of the number of training epochs, the number of neurons in the output LSTM layers, the batch size, the early stopping technique, and the number of heads in the attention layers on the forecast. The purpose of training and using a neural network is to predict the moments in time when the risks of critical events exceed the permissible level, as well as to determine the location of the incident activation taking into account the influencing factors, the combination and values of which can lead to accidents with visualization in the form of a heat map of hazardous areas.

Keywords: Road safety; influencing factors; accident rates; recurrent neural network; transformer; multi-headed attention mechanism; time series; time series forecasting.

---

[*] Correspondence to: Anton Finogeev. E-mail: anton.finogeev.pnzsu@inbox.ru

INTRODUCTION

The causes of road traffic accidents (RTA) are determined by a combination of various factors of influence of subjective and objective nature, which include the state and behavior of road users, road section parameters and operational condition of the road surface, condition of road transport infrastructure (RTI) elements, weather conditions, etc. Subjective factors, such as the behavior and condition of drivers and pedestrians, are quite difficult to detect, model and predict. Objective factors, such as weather conditions and operational condition of roads and RTI elements, can be recorded using sensor devices and processed using intelligent and predictive data analysis technologies. Modern methods make it possible to determine correlations and assess the degree of influence of specific factors or their combination on road incident indicators in order to predict the risks of critical events. The probability of an RTA on road sections is influenced by many parameters that can be expressed in a quantitative assessment of the indicators of influencing factors. However, the analysis and forecast of the indicators of each factor separately does not give very accurate results, since the risks of road accidents are often caused by the complex impact of many external factors in combination with the technical condition of the vehicle, experience and psychophysiological state of drivers and pedestrians before the incident. In this case, the influence of external factors is considered without taking into account the condition of the vehicle and road users. Then the number of incidents Z_i on the i-th road section of length l (km) over a time interval t in days depending on external factors can be defined as a function: $Z_i(t)=P_i(N,C,K,V,S)$, where P_i is the probability of critical situations, N is the indicator of the number of vehicle passages, C is the indicator of vehicle speed, K is the coefficient of adhesion to the road depending on weather conditions (humidity, dew point, temperature, wind strength and direction, depth of water film, type of road surface, height of snow cover, degree of icing, etc.), visibility index B depending on the characteristics of the road section (width of the roadway and shoulder, turning radius, angle of transverse slope, angle of ascent), illumination and weather conditions (fog, rain, snow), S is the index of deformation of the road surface on the section (%) depending on the operational condition of the section (number and size of defects).

During the study, a neural network model, a methodology for training and forecasting critical events in the road transport environment were developed, which is based on the analysis of dynamically changing information on the influence of external factors on the risks of incidents. The prognostic assessment is performed during the analysis of time series of retrospective data on accidents, traffic violations and indicators of objective factors of influence. Such factors include, in particular, characteristics of the road surface condition, traffic indicators, lighting and visibility conditions, and other meteorological data. Data is collected using video surveillance cameras, road traffic sensors, weather road stations, photo radar photo and video recording systems. The results of the study are a technology for prognostic assessment of risks and the time of occurrence of critical events in the road environment in the short and medium term based on a comparative analysis of the predicted time series of incident indicators and factors of influence. The forecast results, namely, the time and place of activation of possible road accidents, are supposed to be displayed on an interactive map with color differentiation of the danger level and its change on the time scale. The solution to the problems of predictive analysis of accident rates and influencing factors will allow notifying road users about dangerous road sections and predicted moments of time of probable occurrence of critical events in order to prevent accidents and improve traffic safety.

Theoretical background

Risk assessment of occurrence and development of critical events in the road transport environment is a necessary element of road safety management. Objective and subjective factors of influence, especially their dangerous combinations that lead to accidents, are the main problem of preventing accidents on road sections. The study (Gatarić *et al.*, 2023) presents neural network models for predicting the number of accidents on roads, the severity and size of material damage depending on objective factors such as the length of a road section, type of relief, width of the roadway, traffic volume and speed limit on the section. The results of predictive analysis are often used in planning and selecting optimal routes (Gamidullaeva & Finogeev, 2023). An important factor, for example, is the operational condition of the road surface, which affects accident rates (Katasonov *et al.*, 2017) in areas with a concentration of accidents. Many methods and models are developed to analyze and model road situations in order to predict accident rates, which are used in practice in various countries. Traditional methods of analysis are statistical models and machine learning methods. Among the statistical research methods, we note the Poisson regression model for studying the relationship between risk factors and accident frequency (Abdulhafedh, 2017; Shaik & Hossain, 2019), the negative binomial regression model (Naghawi, 2018) and the gamma regression model (Shaik & Hossain, 2020). Other forecasting models include the random parametric model (Dinua & Veeraragavan, 2011), the K-nearest neighbors model (Zhang *et al.*, 2013), and the support vector machine (SVM) (Li *et al*., 2008). Machine learning methods such as decision trees, random forest and logistic regression (Vanitha & Swedha, 2023) are also used to predict accidents taking into account weather conditions, vehicle conditions, road surface and lighting. The article (Ayou *et al*., 2023) describes, in particular, machine learning algorithms for incident detection using the support vector machine (SVM), random forest (RF) and a neural network with long short-term memory (LSTM).

In addition to statistical models and machine learning methods, neural network models have recently become popular, which demonstrate flexibility, high performance, sufficient forecast accuracy and the ability to generalize in the process of road traffic studies (Rusul *et al*., 2019; Pradhan & Sameen, 2020). Artificial neural networks can model nonlinear functions to assess safety parameters with the ability to assess the risks of accidents, identify causes, determine the time and place of incidents (Ahmed & Pradhan, 2019; Profillidis & Botzoris, 2019). The problems of intelligent traffic management and DTI components are also solved using the neural network methodology (Ozgan, & Demirci, 2008). Neural network models are used in managing traffic and pedestrian flows, predicting the behavior of road users, modeling the causes of accidents and factors influencing them, to assess the consequences of accidents, the severity of injuries, etc. Modeling road incidents using neural networks helps to assess the indicators of factors associated with technical problems of vehicles, the medical and mental state of drivers, the operational characteristics of the road surface and roadside infrastructure, the influence of weather conditions and other environmental elements. The tasks of forecasting aspects and parameters of road traffic are solved by identifying hidden dependencies and interrelations in sequences of input data and variables describing the state of the road traffic infrastructure and the environment. The authors of the article (Mayura Yeole et al., 2022) present the results of comparing the quality of forecasting using a neural network with forecasts obtained by statistical methods on the same set of input data describing weather conditions, road congestion, parameters of highway sections, time characteristics of incidents, the state of infrastructure elements, etc. In the work (Ogwueleka et al., 2014), the authors evaluate the effectiveness of a neural network in analyzing and forecasting accident rates depending on the number of vehicles on the roads and other parameters in comparison with the traditional statistical approach.

An important task is to develop methods and means for detecting incidents in order to notify road users and relevant services. Currently, automatic incident detection systems are being developed (Tai et al., 2018; Lamr, 2018). A review and analysis of the systems is given in the article (Hireche & Dennai, 2020), which considers studies using machine learning methods. The problem of detecting road accidents is usually presented as a task of classifying road events and determining the degree to which a critical event belongs to an accident (Xie, Shang & Yu, 2022). The article (Bharath Kumar *et al*., 2021) considers a support vector model, which is used to create an effective classifier. In the work (Dogru & Subasi, 2018), a random forest model is used to detect road accidents, which is able to combine uncorrelated decision trees to select a high-quality forecast with better performance than a separate decision tree. The article (Zikratova & Zikratov, 2020) considers the problem of predicting the probability of an accident as a result of traffic violations using a neural network. The neural network model is used to classify situations that arise as a result of traffic violations and to identify dangerous violations that can lead to accidents. The paper (Ki *et al*., 2018) presents a method for detecting traffic accidents depending on the change in the average speed of the traffic flow and speeding violations using a neural network. The paper (Zhu *et al*., 2018) proposes to use a convolutional neural network (CNN) for automatic detection of traffic incidents in the city based on traffic flow data. During the research, the CNN model was compared in efficiency with the multilayer perceptron (MLP) model and showed results with a higher detection rate and a small number of false positives. Technical means are also being developed for traffic accident detection systems, for example, using infrared (Wadhahi et al., 2018) or ultrasonic sensors (Khalil et al., 2018). Note that the detection accuracy depends on traffic parameters and spatio-temporal traffic data (Ghahremannezhad, Shi & Liu, 2022).

Analytical studies show that the most promising forecasting tools are neural network models. In (Ma et al., 2015; Gu et al., 2019), researchers used LSTM deep learning networks for short-term forecasting of the average speed of traffic flows on traffic lanes based on the analysis of data from detectors and transport counters. The article (Alamir, Zargaryan & Zargaryan, 2021) considers a method for forecasting traffic using various neural network models, where the best forecasts were obtained using the multilayer perceptron model. The forecasting results are used to reduce congestion on highways during rush hours. An important task of traffic analysis is to assess the impact on the growth of the probability of accidents on highways (Zhang, Wang & Fang, 2019). The change in the parameters of traffic flows over time is presented here in the form of time series of data. For time series, neural networks are used for (Emaletdinova, Kabirova & Vildanov, 2022; Daub, 2020). Varieties of recurrent neural networks, for example, long short-term memory (LSTM) models (Song et al., 2020), work effectively with time series. In particular, such networks are used for predictive modeling of time indicators of traffic (Polson & Sokolov, 2017).

Neural networks are used to predict the severity of road incidents and their consequences in the form of damage. Predicting the severity of road accidents indirectly affects the reduction of their number. A large number of works are related to the classification and predictive analysis of the severity of road accidents using regression analysis methods and neural networks. The research results show the dependence of the severity of injuries on the speed of the vehicle, the direction of impact during collisions, and the use of seat belts. The article (Labib *et al.*, 2019) presents the results of the analysis of methods for classifying the degree of damage as a result of road accidents. Neural network models for solving this problem provide greater accuracy and efficiency compared to regression analysis methods (Alkheder, Taamneh & Taamneh, 2017; García de Soto *et al.*, 2018; Chakraborty, Mukherjee & Mitra, 2019). In (Abdelwahab & Abdel-Aty, 2002), researchers assessed driver injuries, the amount of material damage to vehicles, road safety facilities, and the environment using a multilayer perceptron model and an adaptive resonance fuzzy network. The article (Olutayo, & Eludire, 2014) compared the results of predicting the severity of injuries using neural networks and decision trees. The authors of (Shaik *et al.*, 2021) also present the results of a comparative analysis of several neural network models in terms of efficiency and performance in predicting severity, including the single-layer perceptron (SLP) model, multilayer perceptron (MLP), radial basis function (RBF) network, recurrent network (RNN) and convolutional network (CNN) models. Similar studies are considered in (Raja *et al.*, 2023), where accident prediction is performed using six architectures, including backward propagation (BPNN) and forward propagation (FFNN) models, MLP, RBF, RNN and LSTM models. Neural networks were used to classify accident types, predict the number of accidents and their severity. The studies show that the best results in terms of prediction accuracy and efficiency were obtained using RNN and CNN network models (Zheng *et al.*, 2019). At the same time, the SLP, MLP, RBF, FFNN models showed sufficient accuracy, and the BPNN model - minimal prediction accuracy. The CNN model for predicting accident severity is discussed in detail in (Zheng *et al.*, 2019). The RNN model showed the best accuracy in predicting the severity of injuries in road traffic accidents in a comparative study with CNN and FFNN models, single-layer (SLP) and multi-layer (MLP) perceptrons (Sameen & Pradhan, 2017).

MATERIALS AND METHODS

The main idea of the proposed approach is not in the prognostic assessment of the risks of road incidents, but in the forecast of the moments in time when the risks of critical events exceed the permissible level, as well as the determination of the locations of possible incidents taking into account the totality of influencing factors, the combination and values of which can lead to an accident with a certain probability. The method for forecasting the time moments of the probable occurrence of critical events works on the basis of data on the indicators of influencing factors. Data are extracted from xml files that are formed in the process of collecting and consolidating information on factors (Finogeev *et al.*, 2021). In the study, indicators of meteorological factors and the operational condition of the road surface were selected for analysis (Finogeev *et al.*, 2024). Time series of changes in these indicators were used to form training and test data sequences.

The forecast results are necessary for alerting and warning road users so that they can take measures to safely drive vehicles in dangerous areas and prevent possible incidents. The works (Finogeev, 2023; Finogeev *et al.*, 2020) consider the methods of assessing possible factors influencing the risks of road incidents and predictive analysis of critical event indicators (Finogeev & Kolesnikov, 2020). In the course of studies based on comparative analysis, similar patterns of correlating series of indicators of influencing factors and indicators of critical events were determined in order to synthesize predictive models for assessing the risks of road accidents. The work (Polezhaev & Finogeev, 2024) proposes a recurrent neural network model with a transformer for predicting time series of critical event indicators in the road transport environment. Predictive modeling was performed for time series of the number of vehicle passages, traffic violations (TRP) and the number of accidents. The input data for the time series were obtained from photo radar photo and video recording systems installed on the roads of Penza and the Penza region. A logical continuation of the research is the analysis of time series of indicators of influencing factors and accidents in order to assess probabilistic risks and moments in time when a combination of factors leads to incidents on specific road sections. The forecasting results should be displayed as a heat map of the predicted danger and time of incidents on road sections. Color differentiation of dangerous and safe sections with the display of the predicted time of incidents is a way of early warning of road users in order to decide on changing the route or travel time, or changing behavior and traffic patterns in predicted time intervals. The results of forecasting the time of exceeding the risks of incidents can also be used in the operation of intelligent unmanned vehicles to automatically change routes and traffic patterns depending on the risks of incidents on road sections.

For prognostic risk assessment and determination of time points, a method for predictive analysis of time series of influencing factor indicators based on a deep learning neural network model is proposed. Analysis of neural network models showed that models based on recurrent or convolutional neural networks, which demonstrate high accuracy and efficiency in predicting road accident indicators, are most suitable for achieving the research goals (Gutierrez-Osorio & Pedraza, 2020; Wenqi, Luo & Yan, 2017). To improve accuracy and performance, the study proposes to use an RNN model with output LSTM layers supplemented by a transformer with multi-head attention mechanisms (Multi-Head Attention) (Vaswani et al., 2017). LSTM layers allow taking into account time sequences by replacing multiple matrix multiplications with weighted sums, which allows better processing and generalization of long data sequences. The proposed model implements three LSTM layers (**Fig. 1**), which include memory cells for storing information and filters that allow using information from long sequences.

Filters include: (a) input filters, which determine which input data is added to the cell state, (b) memory filters, which determine which part of the information will be forgotten, and (c) output filters, which determine which part of the information from the cell state is used to form the output signal.

Since an RNN network with several LSTM layers cannot always process too long time series, it was decided to add a transformer layer with internal attention mechanisms to improve efficiency (Lin et al., 2017). This mechanism allows working with any element of the input data in the process of generating output values, since it establishes relationships between the positions of elements within the sequence and calculates the corresponding vector representation. Integrating a transformer with LSTM layers will allow us to use the transformer model at the data processing stage, and LSTM layers at the final stage. The components of the transformer are:

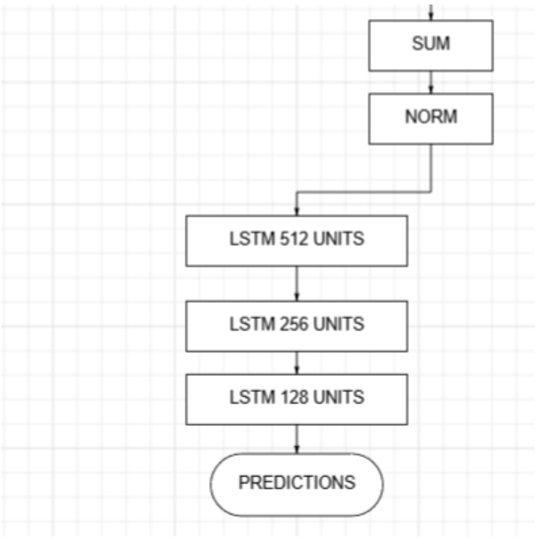

Fig. 1. Output LSTM layers of the model.

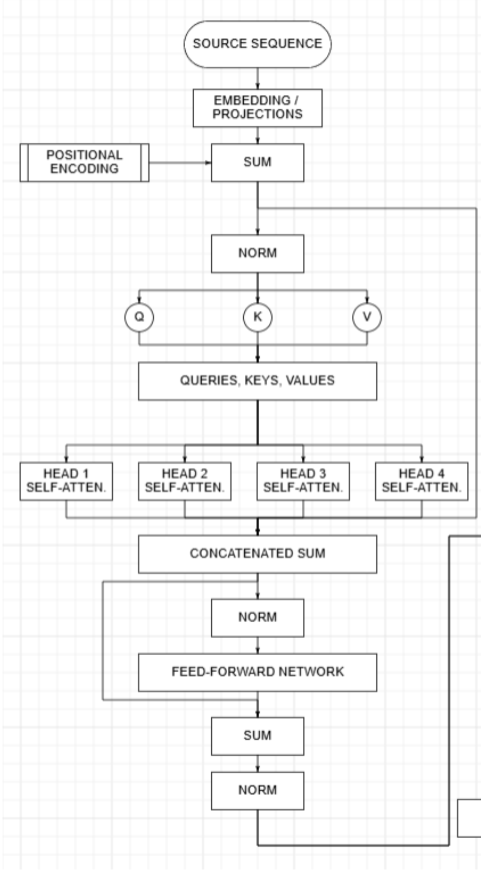

Fig. 2. Multi-head removal mechanism in the transformer encoder.

1. An encoder that includes multiple layers of self-attention and a fully connected network model with feedback from the attention mechanism to process the input sequence and synthesize contextual representations.
2. A decoder, which includes several layers of internal attention with input from the previous layer, a layer of attention to the results of coding with input from the attention mechanism and an encoder layer with input from the attention mechanism.
3. Self-attention layers are formed with multiple heads (**Fig. 2**), which allow the model to work with different parts of the input sequence when forming the output vector. Heads are needed to calculate the weights of a part of the time series elements and use the results when determining the weighted sum.

The input data sequence is fed to the encoder to calculate the output vector representation with the addition of the positional encoding vector. Next, the set of elements, without regard to order, is fed to the attention layer for processing, and then a part of the sequence is fed to the decoder input together with the output vector of the encoder to calculate the resulting vector of the output sequence. Several layers of internal attention can operate in the encoder and decoder (**Fig. 3**).

Each attention layer sequentially passes the processing results to the next layer as input for transformation into context vectors while preserving the context information of each element in the sequence. The decoder receives the context vector representations from the encoder as input and generates the elements of the output vector using a probabilistic model. The decoder layers also pass the processing results to the next layers along with the resulting vectors from the encoder and form a sequence of output vectors. The attention layers work with 3 components: queries, keys, and values. In the decoder, the query is the current state, and the keys and values are the states of the encoder. During operation, the attention layers calculate the similarity function of the query with each key, and then average the resulting values proportionally to the similarity function. Thus, the attention function can be defined as a comparison of a query and a set of key-value pairs with the output data, where the query, keys, values, and output data are vectors. The result is a weighted sum of values, where the weights are calculated according to the similarity of the query and the key. There are two types of attention mechanisms used in neural networks: additive attention (Bahdanau et al., 2014) and multiplicative attention based on the scalar product with scaling (Britz et al., 2017). Our model implements the multiplicative attention mechanism in two layers, which calculates the scalar products of a query of dimension d_k with all keys. Then, each product is divided by $\sqrt{d_k}$ for scaling, and the *softmax* function is used to calculate the weights of the values. The output attention matrix V is calculated from the query matrices Q, keys K, and value matrix Z as:

$$V(Q,K,Z) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)Z \tag{1}$$

The use of the multi-headed internal attention mechanism in the encoder and decoder is used to improve performance by parallel processing of input data vectors in the heads. The mechanism allows working simultaneously with several linear projections of queries, keys and values onto the dimensions d_k, d_q and d_z. On each projection, the attention function is executed and the output values of the dimension d_z are determined, which are then calculated by concatenation:

$$\begin{gathered}(Q,K,Z) = Concat(head_1, \dots, head_h)W^o, \\ \text{where } head_i = V\left(QW_i^Q, KW_i^k, ZW_i^z\right),\end{gathered} \tag{2}$$

where projections are parametric matrices:

$$W_i^Q \in R^{d_m \times d_q},\ W_i^K \in R^{d_m \times d_k},\ W_i^Z \in R^{d_m \times d_z},\ W_i^O \in R^{hd_z \times d_m}. \tag{3}$$

The methodology for software implementation of the model includes the following stages.

1. The input time series data sequence is divided into several streams according to the number of heads.
2. The heads calculate the attention weights for the data stream in parallel.
3. The calculation results from all heads are combined during concatenation.
4. The obtained results are fed to the feedforward layer, which performs linear transformations in the form of convolutions with an activation function between them.

The general scheme of the synthesized neural network model with a transformer, including two layers of multi-headed self-attention in the encoder and decoder, and with three output LSTM layers is shown in **Fig. 4**.

The methodology for synthesizing and training the implementation of the software for its configuration using the example of predicting the number of speeding violations on road sections includes the following stages:

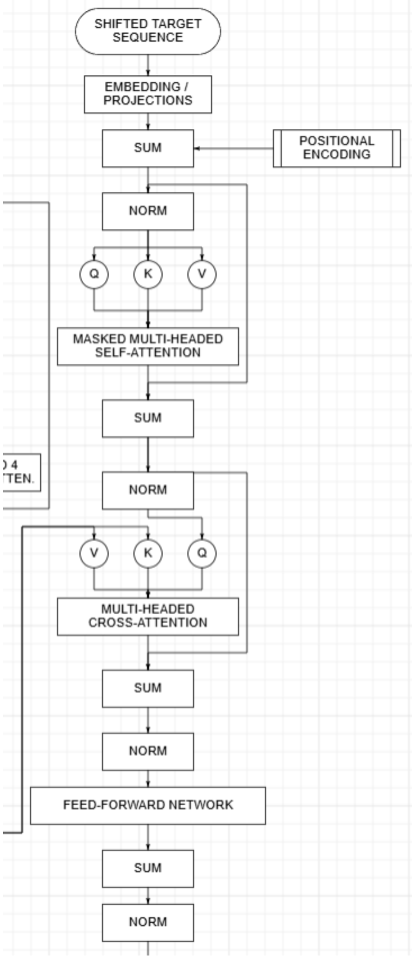

Fig. 3. Encoder layer with two layers of multi-headed attention.

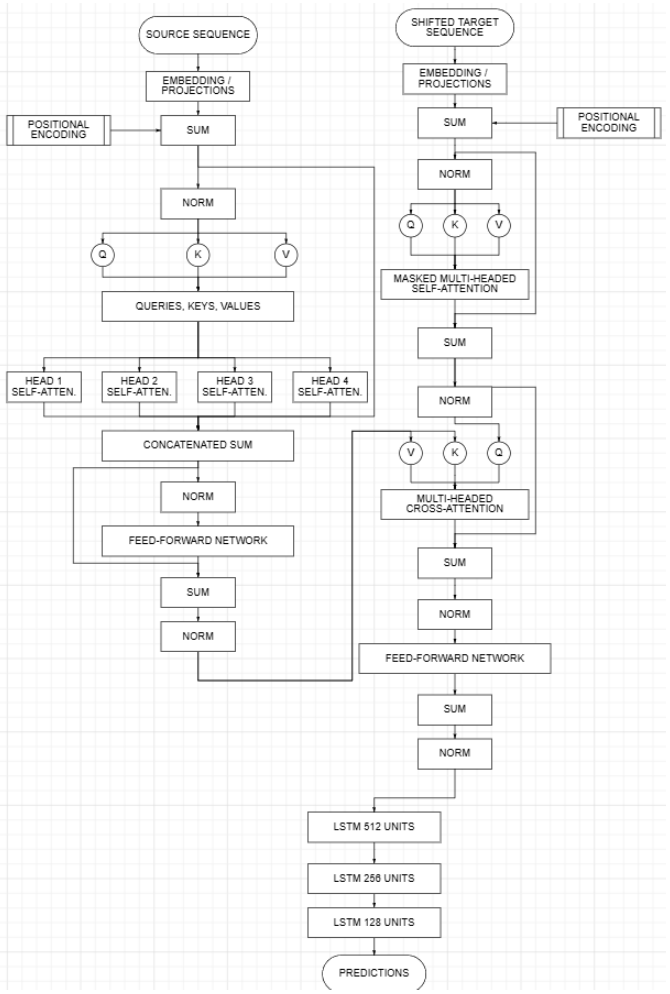

Fig. 4. Neural network with transformer, multi-headed attention and LSTM layers.

1. Input and extraction of data on road accidents and influencing factors from the input Excel file, which is generated as a result of unloading the XML file:

```
sheet_name <- " June 2024"
excel_data <-
readxl::read_xlsx("June2024.xlsx", sheet =
sheet_name)
# Example of speeding violation data
extraction and time series synthesis
speed_data <- excel_data$Speed[1:2000]
time_series <- ts(speed_data, frequency = 1)
```

2. Setting the window size and preparing training and test data for time series modeling

```
prepare_data <- function(series, window_size)
{
  x <- matrix(series, ncol = 1)
  y <- series[-seq_len(window_size)]
  x <- array(x, c(length(y), window_size, 1))
  y <- array(y, c(length(y)))
  list(x = x, y = y)
}
# Determining the size of the window
window_size <- 1000
# Data preparation
data <- prepare_data(speed_data, window_size)
```

3. Synthesis of RNN model with transformer, attention layers and LSTM layers

```
query_input <- layer_input(shape =
c(window_size, 1), name = "query_input")
value_input <- layer_input(shape =
c(window_size, 1), name = "value_input")
# Setting the number of heads and forming a
multi-headed attention layer with
determination of the dimensionality of keys
and values
attention_layer <- layer_multi_head_attention(
  num_heads = 6,  # Number of attention heads
  key_dim = 4     # The size of keys and
values has been increased to 4
)(query_input, value_input)
# Forming LSTM with three output layers and
setting the number of neurons in them
lstm_layer <- layer_lstm(units = 512,
return_sequences = TRUE)(attention_layer)
lstm_layer_2 <- layer_lstm(units = 256,
return_sequences = TRUE)(lstm_layer)
lstm_layer_3 <- layer_lstm(units = 128,
return_sequences = TRUE)(lstm_layer)
output <- layer_flatten()(lstm_layer_3)
output <- layer_dense(units = 1)(output)
model <- keras_model(inputs = c(query_input,
value_input), outputs = output)
```

4. Selecting the batch size and compiling the model for training on a sample that includes 80% of the training time series data and 20% of the test data, with a specified number of training epochs using the early stopping technique

```
model %>% compile(
  loss = "mean_squared_error",
  optimizer = optimizer_adam(learning_rate =
0.001)
)
history <- model %>% fit(
  list(data$x, data$x), data$y,
  epochs = 100,
  batch_size = 10,
  validation_split = 0.2,
  callbacks =
list(callback_early_stopping(patience = 25))
)
```

5. Time series forecasting using a trained neural network model

```
        predictions <- model %>%
    predict(list(data$x, data$x))
      # Forming a data list
        forecast_ts <- ts(predictions,
    frequency = 1)
        cat("The first 10 values of the
    original time series:\n")
        print(head(time_series, 10))
        cat("\nFirst 10 forecast values:\n")
        print(head(predictions, 10))
      # Determining the number of forecast
steps
        forecast_steps <- 1000
        future_predictions <-
    numeric(forecast_steps)
      # Setting the forecast window
        context_size <- 2000
        last_window <- tail(speed_data,
```

```
    window_size)
          long_context <- tail(speed_data,
    context_size)
          for (i in 1:forecast_steps) {
      # Calculating the forecast result at the
forecast step
           next_prediction <- model %>%
    predict(list(array(last_window, dim =
    c(10, window_size, 1)),
    array(long_context, dim = c(10,
    context_size, 1))))
           #next_prediction <- model %>%
    predict(list(last_window, last_window))
            future_predictions[i] <-
    next_prediction
            last_window <- c(tail(last_window,
    window_size - 1), next_prediction)
            #print(tail(last_window, 5))
            long_context <-
    c(tail(long_context, context_size - 1),
    next_prediction)
          }
# Synthesis of a new time series with forecast
results
extended_time_series <- c(speed_data,
future_predictions)
extended_forecast_ts <-
ts(extended_time_series, frequency = 1)
```

6. Visualization of time series forecasting results

```
plot(time_series, type = "l", col = "blue",
lwd = 2, ylim = c(0,
max(extended_time_series)), xlim = c(1,
length(extended_time_series)))
lines(ts(extended_time_series, frequency = 1),
col = "#00ff40", pch = 16)
lines(ts(c(speed_data, future_predictions),
frequency = 1), col = "red", pch = 16)
lines(forecast_ts, col = "#b3ff00", pch =
16)
#lines(c(rep(NA, window_size), forecast_ts),
col = "#b3ff00", pch = 16)
legend("topright", legend = c("Initial
series", "Extended forecast"), col = c("blue",
"red"), lty = 1, lwd = 2)
```

Preparation and study of time series parameters is implemented in Python in the R statistical computing software environment. For training, testing and forecasting, the neural network model is synthesized using the TensorFlow deep learning framework, Keras libraries (a high-level interface for creating and training neural networks), Tokenizer and pad_sequences for processing input data, as well as the Gensimc TensorFlow library. The Keras library allows you to create complex neural network architectures with LSTM layers, a transformer and attention mechanisms.

Results of the study of the structure and parameters of the neural network

Before using the model, it is necessary to conduct its research in the process of training and testing to assess the impact of different options and parameters on the accuracy and efficiency of forecasting. The result is the adjustment of its structure and parameters to obtain an acceptable forecast error. For training, it is necessary to select training and test data sets. In the project, data sets are formed based on the results of collecting accident indicators and influencing factors in the road environment. As an example, we will use time series of changes in the number of accidents and the vehicle speed indicator on a section of the road as one of the influencing factors. According to statistics, about a third of accidents occur due to drivers failing to comply with speed limits (Maiorov, 2024). Speeding also leads to an increase in traumatic accidents. According to the World Health Organization, an increase in vehicle speed by 1 km / h leads to an increase in such accidents by 3%, and fatal accidents by 4-5% (WHO report “Speed Management”, 2017). Data for the synthesis of time series were obtained from photo radar systems for photo and video recording of road incidents and traffic violations. The information is related to big data, since the collection process records a huge number of vehicle passages, traffic violations, road incidents and changes in other factors of influence on many road sections. For example, in June 2024 alone, 40,843 vehicles with 6,856 speeding violations, 367 accidents, including 56 accidents with injuries, were recorded on 6 road sections in the Penza region. An example of the distribution of vehicle speed indicators on 18 road sections in the city of Penza and the region is shown in **Fig. 5**.

To train the model in order to select its optimal structure and parameters, the input time series data set is divided into two parts. Two division methods were used in the study: a) 50% training and 50% testing parts, b) 80% training and 20% testing parts. The first part of the series is used to train the model, and the second part is used to compare the predicted values with the actual ones and estimate the error. The goal is to minimize the error σ, which is tested on the test set $\{(\mathrm{Xj}\ , \mathrm{Yj}\ )\}_{j=1}^{J}$, where X_j is the time series data, and Y_j is the predicted value. After estimating the error, the forecast model is adjusted and retrained in an iterative mode. To terminate the iterative process, the early stopping technique or the moment of reaching an acceptable level of error is used. Let us consider the process of studying the variants and parameters of a neural network model with the aim of optimizing it on training sets before the forecasting stage.

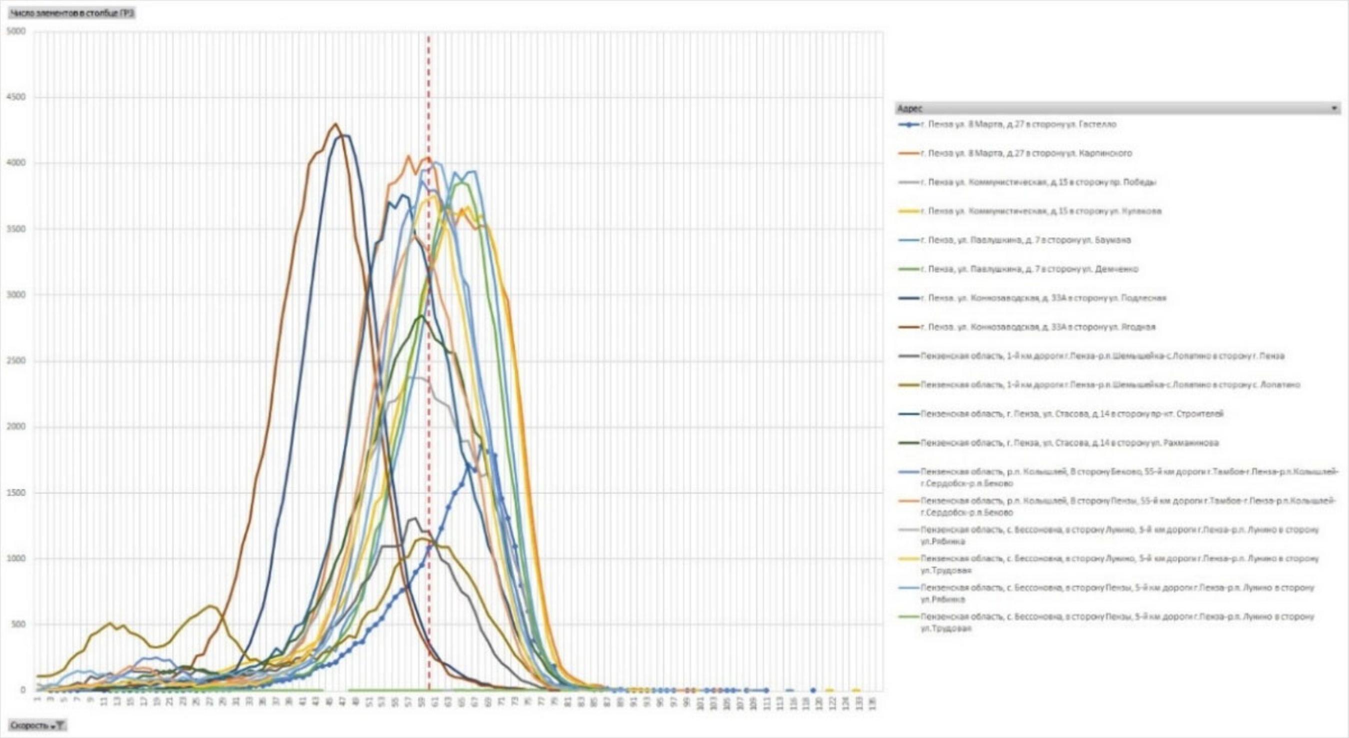

Fig. 5 Distribution of the number of vehicles with speed indicators on 18 sections.

Model study to determine the number of training epochs

To train the model, it is necessary to set the number of training epochs so as to ensure a minimum forecast error and the ability to generalize time series of data. A training epoch is one run of the training data set through the neural network. During the training process, the time series data passes through the model for a given number of epochs, and the network weights are updated based on the errors calculated during forecasting on the test part of the data set. The results of model studies for different numbers of epochs show how their number affects the quality of model training. With a small number of training epochs, the model may not have time to learn on a given amount of data, and a situation of underfitting occurs. As a result, the model will not be able to accurately and reliably identify hidden dependencies in the data series, which will lead to a large forecast error on the test set. On the other hand, too many epochs can lead to a situation of overfitting, when the model begins to accurately predict the results on the training set, but will show a large error on the test and real samples. A similar situation may arise when using complex models with a large number of neurons, or when working with input datasets with repetitive patterns (Chip Huyen, 2022). To achieve a balance between underfitting and overfitting, it is necessary to select the optimal number of epochs that will provide the required accuracy and generalization ability. The optimal number depends on the following characteristics (Lakshmanan, 2022):

1. The size and complexity of the training dataset. Larger datasets require more epochs for training.
2. Structures of the neural network model. Models with a complex and multi-level structure require more time for training.
3. Method of stopping the training process. In the process of training a model, the early stopping technique is often used, when the training process stops if the forecast error on the test data set stops decreasing over a specified number of epochs.

Model studies were conducted on a neural network including 2000 neurons in three output LSTM layers. The number of epochs was set from 10 to 200 with a step of 10. For the number of epochs less than 60, a situation of undertraining of the model was observed, and for more than 160 epochs, a situation of overtraining. The graphs show the forecasting results for 40, 60 and 150 epochs of neural network training (**Fig. 6**). The optimal number of epochs for training our model was set to 100.

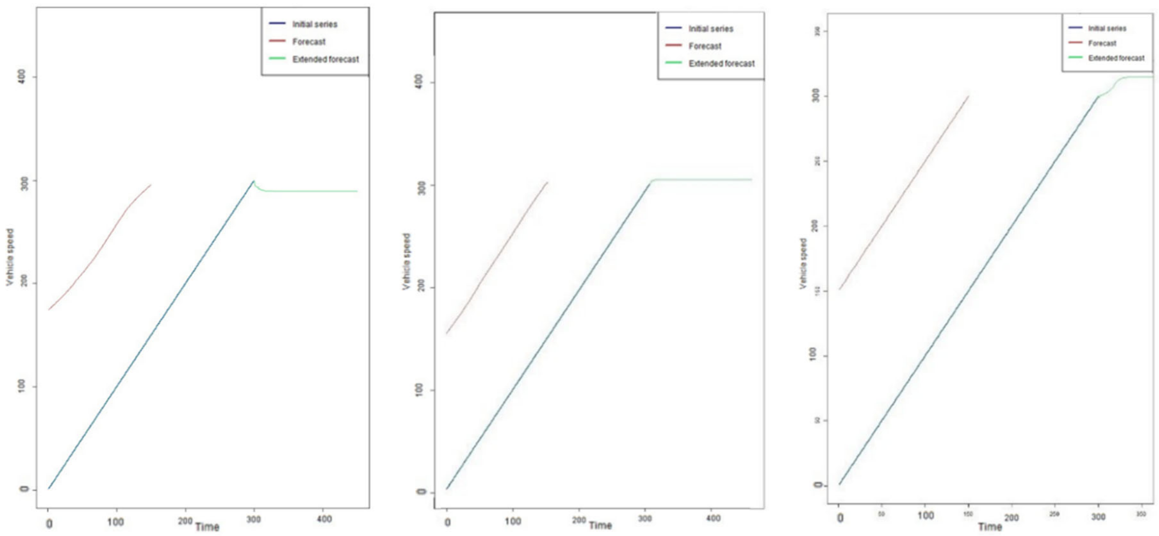

Fig. 6 Simulation results for different numbers of training epochs.

Modeling studies for choosing the number of neurons in LSTM layers

The next parameter of the model that affects the forecast results and can create situations of underfitting and overfitting is the number of neurons in the output LSTM layers. The number of neurons here affects the learning ability when analyzing complex and large data sequences. A large number of neurons allows the model to identify complex and long-term dependencies. A small number of neurons is used for simple tasks with short and less complex dependencies. With an increase in the number of neurons in the model layers, the number of its parameters increases, as well as the computational complexity. The training process slows down, and the inference time increases. Since inference is a practical stage of applying the model to real data after training and testing, the goal of optimization is to achieve high performance and minimal delays when using it on specific types of devices (Gusak et al., 2021).

For a neural network model, in each case it is necessary to select the optimal number of neurons in LSTM layers, since an insufficient number will not allow identifying the necessary dependencies, and too large a number can lead to a state when the model remembers the training data, but generalizes poorly in the case of new data. In the process of studying the model, experiments were conducted for models with different numbers of neurons in LSTM layers. As an example, the results of studies are given: a) a model with a small number of neurons, where 128 neurons are specified in the first layer and 64 in the other two, b) a large model with the same number of neurons, when 2000 neurons are specified in all layers (**Fig. 7**). The number of training epochs was taken to be 100. Based on the research results, an intermediate model with 512 neurons in the first layer, 256 neurons in the second layer and 128 neurons in the third layer was synthesized.

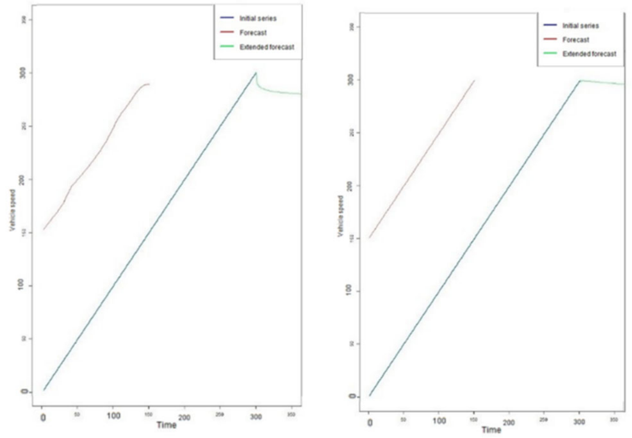

Fig. 7 Simulation results for small and large models.

Modeling studies for batch size selection

A recurrent neural network with a transformer and attention mechanisms requires a large memory and computational resources, since it is necessary to store all the keys and values (internal states of the encoder). The computational complexity of determining the context vector when generating output data is directly proportional to the amount of data in the input time series. To reduce the training time of the model, the training data set is divided into subsets (batches), which are processed separately. In this case, the model updates the weights after running each batch, rather than the entire training set. Batches are used in the case of large data sets, so as not to run the entire set through the network when recalculating the weights. Dividing the training set into batches leads to an acceleration of the model training process and is called batch normalization (Ioffe & Szegedy, 2015). Therefore, an important parameter is the batch size, which affects:

1. Convergence and stability of training. When batch size is small, weights are updated frequently, which can lead to a noisy gradient estimate. This can help avoid local minima, but it makes convergence more difficult. When batch size is large, gradient descent is more stable and accurate in optimization, but it requires more training time.
2. Computational efficiency and memory usage. Small batches require less memory and computational power, but may make the network model run slower on modern architectures. Large batches require more memory, but allow using different heads of the attention mechanism for parallel data processing.
3. Generalization ability. Small batches can contribute to better generalization ability and help avoid overfitting. Large batches can lead to overfitting because the model gets more accurate gradients.

As an example of research, the graphs below show the results of training the model based on 500 values of the training set. In the first case, the set is divided into 5 batches of 100 values, and in the second - into 20 batches of 25 values. The results obtained of training the model are shown for 100 training epochs (**Fig. 8**).

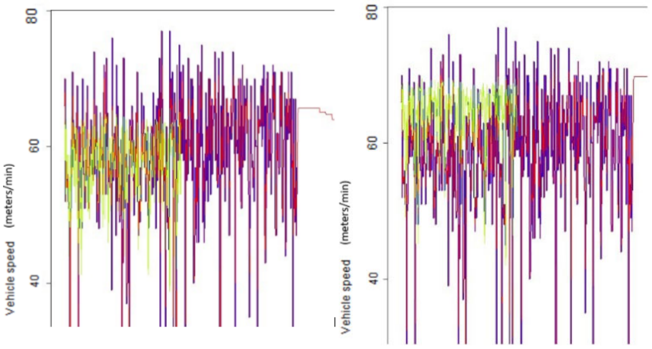

Fig. 8 Simulation results for 5 and 20 batches.

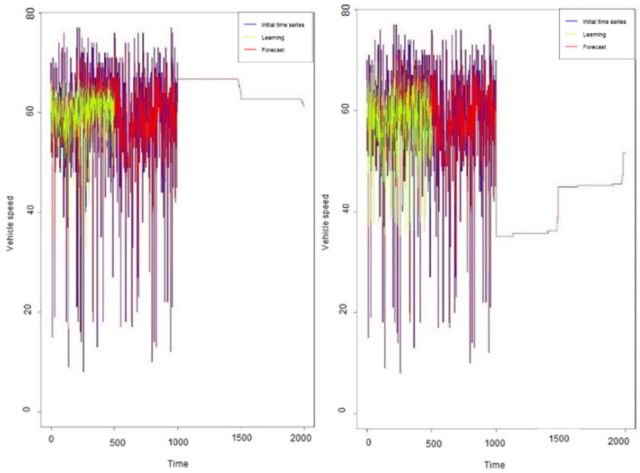

Fig. 9 Simulation results with and without the early stopping technique.

Model studies of early stopping techniques

To forecast complex time series, the model uses a multi-headed attention layer, which provides higher forecast accuracy but increases the training time. Training is usually performed until the error reaches a specified value, which may take a long time and may not lead to the desired result. Therefore, the early stopping technique is used as a mechanism to limit the training time of the model. In this case, the training process stops when the forecast error on the test set stops decreasing over a specified number of epochs. For the purposes of our study, the dataset was divided into training and testing parts in a ratio of 80% to 20%, where 80% of the series is used as training data and 20% as test data. Previously, the batch size was set to 100 values, and the number of training epochs was selected experimentally for training without early stopping and for training using the early stopping technique (**Fig. 9**).

Modeling studies for selecting the number of heads in attention layers

The key component of the transformer is the self-attention layers in the encoder and decoder, which are used to process the input data. In our case, multi-headed attention layers are used, where the heads process parts of the input data in parallel and independently with different weight matrices, which improves the generalization ability and increases the model performance for large data sets. The calculation results in the encoder and decoder of the transformer are combined (concatenated) and passed to the output LSTM layers to obtain the final result. In this way, the model works with different levels of data abstraction, which increases its generalization ability and reveals hidden dependencies. For the multi-headed attention layer to work, it is necessary to select the number of heads in the layers with an estimate of the time complexity and forecast error. Examples of modeling for attention layers with different numbers of heads are shown in **Fig. 10**.

The number of heads in the attention layers determines how many embeddings the model will process simultaneously. With a small number of heads, the model has a less complex structure, which may not be enough to predict very large time series. Using a large number of heads allows the model to analyze various aspects of the data, which improves the ability to identify hidden dependencies and generalize. However, this increases the computational complexity and requires more memory for the model to run. For our project, a model with 6 heads for all attention layers in the encoder and decoder was chosen.

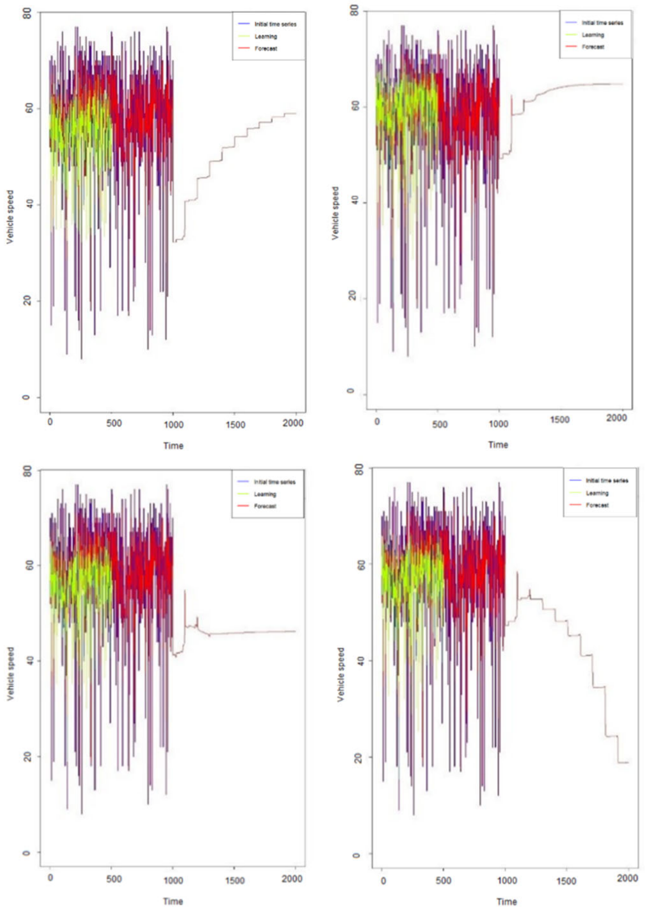

Fig. 10 Simulation results for 2,4,6 and 8 heads in the encoder and decoder layers.

Discussion

During the research, the structure of the neural network model for forecasting the risks and time points of critical incidents was developed based on the analysis of the spectrum of time series of indicators of influencing factors. Similar problems are solved by researchers in our country in order to improve road safety (Mayorov, 2018; Sinitsyn et al., 2022). The system process of architecture synthesis and optimization of the neural network for forecasting the risk of accidents and the frequency of accidents, in particular, are presented in the work (García de Soto et al., 2018). In our case, design studies were also carried out to optimize the structure of the neural network and select its training parameters. The model is a recurrent neural network with a transformer, including an encoder and a decoder with two multi-headed layers of internal attention, and three output LSTM layers for final calculations. The model is developed to assess and predict accident rates considering a set of objective factors such as traffic violations, weather conditions, road conditions and traffic flow parameters. The article (Bulatov, 2022) defines objective and subjective factors of accident risk and proposes a methodology for assessing the impact of their combination on safety. Most objective factors of influence can be assessed quantitatively, and the dynamics of their change can be presented in the form of time series for the purpose of forecasting, which is difficult in the case of subjective factors associated with the behavior and condition of road users.

Modern research shows that recurrent neural network models are functional and effective in predicting accident and disaster risks in real time (Yuan et al., 2019). Before using our model for real forecasting, experimental studies were conducted during its training and testing to assess the impact of its structure and parameter options on time costs, accuracy and efficiency. The result is the choice of the model structure and the adjustment of the training parameters to obtain an acceptable error and forecast time. The main difference in the results of our studies is that the neural network model is used not so much for predicting the risk of accidents, accident rates or severity, as shown in most similar works, but mainly for determining the time intervals before the expected occurrence of critical events. Such moments are recorded when the values of the influencing factor indicators in the predicted part of the time series, in aggregate or separately, lead to a sharp increase in accident rates or the probability of an accident on a specific section of the road. Determining the time intervals preceding the occurrence of a probable critical event will make it possible to create early information alerts for drivers to prevent incidents by increasing their attentiveness and caution. Road sections where the risks of negative changes in factor indicators are identified during the forecasting process are marked in color as possible areas of accident concentration, which is also used to warn road users.

The main problem is the complexity of the problem solution, since for each road section in the spatial zone it is necessary to form and analyze many time series for the prognostic assessment of accident rates taking into account traffic violations and other indicators of influencing factors. In addition, accident rates can be influenced not only by individual factors, but also by their combinations. Therefore, the complexity of solving the problem increases significantly, since in the process of training and validating the model it is necessary to synthesize a model capable of generalizing long time series at different levels of abstraction and identifying hidden correlation dependencies of accident rates and combinations of critical values of indicators of various influencing factors. Thus, training, testing and predictive modeling should be carried out simultaneously on the entire spectrum of time series. At the same time, several factors, such as weather conditions, can affect all road sections in the selected zone, and the results of their forecast can be used for all sections. However, there are factors that affect accident rates only in specific places, for example, section parameters (slopes, curvature, turns, road surface condition, etc.), which requires an impact assessment for each section separately. It is also necessary to consider the correlation between traffic violations and influencing factors, when drivers violate the rules due to the influence of subjective or objective factors. For example, the parameters of a road section related to its straightness, absence of turns and slopes, good visibility and excellent weather conditions often contribute to violation of the speed limit, decreased attention of drivers, which leads to an increase in the risk of accidents instead of increasing safety. Therefore, the task of the research is to identify hidden connections between the risks of accidents, accident rates, violations and other influencing factors through neural network modeling of time series. Based on the analysis of predicted changes in accident rates, violations and other factors, it is necessary to determine the time intervals until the moment when the risks of possible road incidents exceed permissible values, which is the main goal of the research.

A review of studies related to solving similar problems of predicting the time characteristics of accidents is presented in (Tang et al., 2020). However, it should be noted that almost all studies are aimed at predicting the duration of an accident, and not at determining the time intervals before the critical event. For example, in (Zhu *et al*., 2021), a hybrid neural network model consisting of a multilayer perceptron (MLP) and an LSTM layer is proposed for dynamically predicting the duration of an incident. Just like in our project, the forecasting technique takes into account a number of objective influencing factors and traffic flow parameters in real time. Similar studies in the field of predicting the duration of critical events using a neural network model are presented in (Wei & Lee, 2007). The forecast results are used to estimate the size and downtime in traffic jams caused by accidents, as well as to determine the time to clear a section of the road from congestion in order to notify road users.

CONCLUSION

The hybrid neural network model and the methodology of its structural and parametric synthesis are designed to predict the temporal and spatial parameters of road incidents. The forecast results will allow predicting the time and place of possible accidents. Road sections with increased risks of road incidents and time intervals before the probable activation of a critical event are supposed to be visualized on the map in order to warn drivers not only about incidents and traffic jams that have occurred, as is the case in modern navigation systems, but also about probable accidents in the future. Such warnings are proposed to be implemented by dynamically coloring road sections in shades of green (no risk of incidents), yellow (warning of a low probability of an incident) and red (high probability of an incident). In this case, changing the red color to darker and brighter shades means an increase in the degree of danger and a decrease in the time before the probable occurrence of a critical event. Visualization of possible incidents on the map with dynamic change of color designations of dangerous areas in previous time intervals will be used for early warning of drivers, which will increase road safety through their more careful behavior in places with increased risks of critical events. Even not entirely accurate forecast results will reduce the likelihood of road incidents and minimize the impact of objective factors on accident rates.

Another effect is the identification of potential accident concentration areas depending on the impact of a combination of factor indicators in predicted time intervals, which will enable road services to take appropriate measures and implement activities to improve traffic safety in areas of potential accidents. Examples of activities include the installation of warning and restrictive road signs, information boards, application or renewal of road markings, intelligent adjustment of traffic lights, increased speed control, as well as other traffic violations, repair and restoration of the operational condition of the road surface, etc. The results of predictive analysis of the time of deterioration of traffic conditions in specific areas will reduce the impact of objective factors, the impact of which does not depend on the quality and efficiency of the road transport services. Cartographic notification of drivers about the time and place of activation of possible incidents will help prevent accidents and reduce injury rates. The neural network model, the methodology for its training, configuration and predictive modeling are implemented on the platform of an intelligent critical event monitoring system.

ACKNOWLEDGMENTS The research was carried out with the support of the Russian Science Foundation as part of the project № 20-71-10087.

REFERENCES

Abdelwahab, H.T., & Abdel-Aty, M.A. (2002). Artificial Neural Networks and Logit Models for Traffic Safety Analysis of Toll Plazas. *Transportation Research Record, 1784*, 115 - 125. DOI: 10.3141/1784-15

Abdulhafedh, A. (2017). Road crash prediction models: different statistical modeling approaches. *Journal of Transportation Technologies,* 7(2), 190–205. DOI: 10.4236/jtts.2017.72014

Ahmed, A.A. & Pradhan, B. (2019). Vehicular traffic noise prediction and propagation modelling using neural networks and geospatial information system. *Environmental Monitoring and Assessment,* 191(3), 190. DOI: 10.1007/s10661-019-7333-3

Alamir, H. S., Zargaryan, E. V. & Zargaryan, Yu. A. (2021). A *model for predicting traffic flow based on neural networks for predicting traffic on roads. Bulletin of SFedU. Technical sciences,* 6(223), 124-

132. DOI 10.18522/2311-3103-2021-6-124-132

Alkheder, S., Taamneh, M. & Taamneh, S. (2017). Severity prediction of traffic accident using an artificial neural network. *Journal of Forecasting*, 36(1), 100–108. DOI: 10.1002/FOR.2425

Ayou, M., Trardi, Y., Chakir, L. & Boumhidi, J. (2023). Data-driven traffic incident detection in urban roads based on machine learning algorithms. *E3S Web of Conferences*, 469(2). DOI: 10.1051/e3sconf/202346900102.

Bahdanau, D., Cho, K. & Bengio, Y. (2014). *Neural Machine Translation by Jointly Learning to Align and Translate.* ArXiv. 1409.

Bharath Kumar, M., Basit, A., Kiruba, M.B., Giridharan, R. & Keerthana, S.M. (2021). Road Accident Detection Using Machine Learning. In *Proceedings of the 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), Puducherry, India,* pp. 2–6. DOI**:** 10.1109/ICSCAN53069.2021.9526546

Britz, D., Goldie, A., Luong, M. & Le, Q. (2017). *Massive Exploration of Neural Machine Translation Architectures.* 1442-1451. DOI: 10.18653/v1/D17-1151.

Bulatov, S.V. (2022). Analysis of factors influencing the likelihood of road traffic accidents. *Transport engineering*, 6(6), 42-47. DOI: 10.30987/2782-5957-2022-6-42-47

Chakraborty, A., Mukherjee, D., & Mitra, S. (2019). Development of pedestrian crash prediction model for a developing country using artificial neural network. *International journal of injury control and safety promotion*, *26*(3), 283–293. https://doi.org/10.1080/17457300.2019.1627463

Chip Huyen (2022). *Designing Machine Learning Systems*. Sebastopol: O'Reilly Media, Inc., 206 p.

Daub, I.S. (2020). *Review of methods for forecasting time series using artificial neural networks, student,* 3(10), 247–257.

Dinua, R.R. & Veeraragavan, A. (2011). Random parameter models for accident prediction on two-lane undivided highways in India. *Journal of Safety Research*, 42(1), 39-42. DOI: 10.1016/j.jsr.2010.11.007

Dogru, N. & Subasi, A. (2018). Traffic Accident Detection Using Random Forest Classifier. In *Proceedings of the 2018 15th Learning and Technology Conference (L&T),* Jeddah, Saudi Arabia, pp. 40–45. DOI**:** 10.1109/LT.2018.8368509

Emaletdinova, L.Yu., Kabirova, A.N. & Vildanov, N.R. (2022). Neural network modeling of the problem of forecasting time series values. *Bulletin of the Technological University,* 25(10), 101-106. DOI: 10.55421/1998-7072_2022_25_10_101

Finogeev, A. A. (2023). Intelligent analysis of sensory data based on fuzzy logic and neural network in the critical events monitoring system. *Models, systems, networks in economics, technology, nature and society,* 3, 125–141. DOI: 10.21685/2227-8486-2023-3-9.

Finogeev, A., Deev, M., Finogeev, A. & Kolesnikoff, I. (2020). Proactive Big Data Analysis for Traffic Accident Prediction. *Proceeding of the 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA),* 25-27. DOI: 10.1109/CITISIA50690.2020.9371796.

Finogeev, A., Deev, M., Finogeev, A. & Parygin, D. (2024). Recognition and Clustering of Road Pavement Defects by Deep Machine Learning Methods. In: Silhavy, R., Silhavy, P. (eds) *Machine Learning Methods in Systems. CSOC 2024. L52ecture Notes in Networks and Systems,* vol 1126. Springer, Cham. pp. 390-402. https://doi.org/10.1007/978-3-031-70595-3_4807/978-3-030-87034-8_25.

Finogeev, A., Parygin, D., Schevchenko, S., Finogeev, A. & Ather D. (2021). Collection and Consolidation of Big Data for Proactive Monitoring of Critical Events at Infrastructure Facilities in an Urban Environment. In: Kravets A.G., Shcherbakov M., Parygin D., Groumpos P.P. (eds) *Creativity in Intelligent Technologies and Data Science. CIT&DS 2021. Communications in Computer and Information Science*, vol. 1448. DOI: 10.1007/978-3-030-87034-8_25

Finogeev, A.G. & Kolesnikov, I.N. (2020). Proactive monitoring of events based on predictive analysis of time series. *Models, systems, networks in economics, engineering*, 1(33), 111-121. DOI 10.21685/2227-8486-2020-1-9.

Gamidullaeva, L. A. & Finogeev, A. G. (2023). Methodological approaches to managing the development of sectoral ecosystems (on the example of the tourism industry). *П-Economy,* 16(2), 7–23. https://doi.org/10.18721/JE.16201.

García de Soto, B., Bumbacher, A., Deublein, M. & Adey, B.T. (2018). Predicting road traffic accidents using artificial neural network models. *Infrastructure Asset Management*. 5(4), 132–144, https://doi.org/10.1680/jinam.17.00028

Gatarić, D., Ruskic, N., Aleksić, B., Đurić, T., Pezo, L., Lončar, B. & Pezo, M. (2023). Predicting Road Traffic Accidents-Artificial Neural Network Approach. *Algorithms*, 16, 257. DOI: 10.3390/a16050257.

Ghahremannezhad, H., Shi, H. & Liu, C. (2022). *Real-Time Accident Detection in Traffic Surveillance Using Deep Learning.* DOI*:* 10.48550/arXiv.2208.06461.

Gu, Y., Lu, W., Qin, L., Li, M. & Shao, Z. (2019), Short-term prediction of lane-level traffic speeds: a fusion deep learning model. *Transportation Research Part C: emerging Technologies,* 106, 1-16. DOI: 10.1016/j.trc.2019.07.003

Gusak, J., Daulbaev, T., Ponomarev, E., Cichocki, A. & Oseledets, I. (2021). Reduced-Order Modeling of Deep Neural Networks. *Computational Mathematics and Mathematical Physics*, 61, 774–785. https://doi.org/10.1134/S0965542521050109

Gutierrez-Osorio, C. & Pedraza, C. (2020). Modern data sources and techniques for analysis and forecast of road accidents: a review. *Journal of Traffic and Transportation Engineering (English Edition),* 7(4), 432-446. DOI: 10.1016/j.jtte.2020.05.002

Hireche, S. & Dennai, A. (2020). Machine Learning Techniques for Road Traffic Automatic Incident Detection Systems: A Review. In: Hatti, M. (eds) *Smart Energy Empowerment in Smart and Resilient Cities. ICAIRES 2019. Lecture Notes in Networks and Systems*, 102. Springer, Cham. https://doi.org/10.1007/978-3-030-37207-1_7.

Ioffe, S. & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *International conference on machine learning*, 448-456.

Katasonov, M.V., Leskin, A.I., Kochetkov, A.V., Syroezhkina, M.A., Schegoleva, N.V. & Zadvornov, V.Yu. (2017). Mathematical model for predicting road accidents on the road network and in places of concentration of road accidents. *Internet journal "NAUKOVEDENIE",* 9(1). Retrieved from: http://naukovedenie.ru/PDF/33TVN117.pdf.

Khalil, U., Nasir, A., Khan, S., Ali, T., Raza, S.A. & Siddiqui, A. (2018). Automatic Road Accident Detection Using Ultrasonic Sensor. *2018 IEEE 21st International Multi-Topic Conference (INMIC),* 206-212. DOI: 10.1109/INMIC.2018.8595541.

Ki, Yong-Kul & Kim, Joong-Hyo & Kim, Tae-Kyoung & Heo, Nak-Won & Choi, Jin-Wook & Jeong, Jun-Ha. (2018). Method for Automatic Detection of Traffic Incidents Using Neural Networks and Traffic Data. *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON),* 184-188. DOI: 10.1109/IEMCON.2018.8615081.

Labib, F., Rifat, A., Hossain, M., Das, A. & Nawrine, F. (2019). Road Accident Analysis and Prediction of Accident Severity by Using Machine Learning in Bangladesh. *7th International Conference on Smart Computing & Communications (ICSCC),* 1-5. DOI: 10.1109/ICSCC.2019.8843640.

Lakshmanan, V. (2022). *Machine Learning. Design Patterns*. St. Petersburg: BHV-Petersburg, 448 p.

Lamr, M. (2018). Big Data and Its Usage in Systems of Early Warning of Traffic Accident Risks. *2018 Sixth International Conference on Enterprise Systems*, pp. 154-157. DOI: 10.1109/ES.2018.00031

Li, X., Lord, D., Zhang, Y., & Xie, Y. (2008). Predicting motor vehicle crashes using Support Vector Machine models. *Accident; analysis and prevention*, *40*(4), 1611–1618. https://doi.org/10.1016/j.aap.2008.04.010

Lin, Z., Feng, M., Dos Santos, C., Yu, M., Xiang, B., Zhou, B. & Bengio, Y. (2017). *A Structured Self-attentive Sentence Embedding.* DOI: 10.48550/arXiv.1703.03130.

Ma, X., Tao, Z., Wang, Y., Yu, H. & Wang, Y. (2015), Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: emerging Technologies,* 54, 187-197. DOI: 10.1016/J.TRC.2015.03.014

Maiorov, V. I. (2024). Exceeding the speed limit of a vehicle as a risk factor for road traffic accidents. *Modern Science*, 1, 24-27.

Mayorov, V.I. (2018). Risk management in the road safety system. *Bulletin of the Ural Institute of Economics, Management and Law,* 3(44), 8-12.

Mayura Yeole, Rakesh Kumar Jain & Radhika Menon (2022). Prediction of Road Accident Using Artificial Neural Network. *International Journal of Engineering Trends and Technology,* 70(3), 151-161. https://doi.org/10.14445/22315381/IJETT-V70I2P217.

Naghawi, H. (2018). Negative binomial regression model for road crash severity prediction. *Modern Applied Science*, 12(4), 38. DOI: 10.5539/mas.v12n4p38

Nischitha, M., Pavankumar, J., Sheethal, J., Shreya, S. & Suman, J. (2022). Accident Severity Prediction using Machine Learning. *International Journal of Advanced Research in Science, Communication and Technology*, 858-862. DOI: 10.48175/IJARSCT-5172.

Ogwueleka, F., Misra, S., Ogwueleka, T. & Fernandez-Sanz, L. (2014). An Artificial Neural Network Model for Road Accident Prediction: A Case Study of a Developing Country. *Acta Polytechnica Hungarica*, 11, 177-197.

Olutayo, V.A. & Eludire, A.A. (2014). Traffic Accident Analysis Using Decision Trees and Neural Networks. *International Journal of Information Technology and Computer Science (IJITCS),* 6(2), 22-28. DOI:10.5815/ijitcs.2014.02.03

Ozgan, E. & Demirci, R. (2008). Neural Networks-based Modeling of Traffic Accidents in Interurban Rural Highways, Duzce sampling. *Journal of Applied Sciences,* 8, 146-151. DOI**:** 10.3923/jas.2008.146.151

Polezhaev, M.N. & Finogeev, A.A. (2024). Predictive analysis of critical event indicators using a recurrent neural network with a transformer. *Modern science-intensive technologies,* 4, 63-68; https://doi.org/10.17513/snt.39974

Polson, N.G. & Sokolov, V.O. (2017). Deep learning for short-term traffic flow prediction. *Transportation Research Part C: emerging Technologies*, 79, 1-17. DOI: 10.1016/j.trc.2017.02.024

Pradhan, B. & Sameen, M.I. (2020). Review of traffic accident predictions with neural networks. In: *Laser Scanning Systems in Highway and Safety Assessment, Technology & Innovation (IEREK Interdisciplinary Series for Sustainable Development).* Springer, Cham, pp. 97–109. DOI: 10.1007/978-3-030-10374-3_8

Profillidis, V.A. & Botzoris, G.N. (2019). Artificial intelligence—neural network methods. In: *Modeling of Transport Demand Analyzing, Calculating, and Forecasting Transport Demand*, pp. 353–382. DOI: 10.1016/B978-0-12-811513-8.00008-X

Raja, K., Kaliyaperumal, K., Velmurugan, L. & Thanappan, S. (2023). Forecasting road traffic accident using deep artificial neural network approach in case of Oromia Special Zone. *Soft Computing*, 27, 16179–16199. https://doi.org/10.1007/s00500-023-08001-6

Rusul, A., Hussein, D., Sohani, L. & Saeed, A.B. (2019). Applications of artificial intelligence in transport: an overview. *Sustainability*, 11(1), 189. DOI: 10.3390/su11010189

Sameen, M.I. & Pradhan, B. (2017). Severity prediction of traffic accidents with recurrent neural networks. *Applies Sciences*, 7(6)476. DOI: 10.3390/app7060476.

Shaik, M.E. & Hossain, Q.S. (2019). Accident prediction by using Poisson regression for unsignalised junction in Khulna Metropolitan City, Bangladesh. In: *Proceedings of International Conference on Planning, Architecture and Civil Engineering,* RUET, ICPACE-2019, 2019.

Shaik, M.E. & Hossain, Q.S. (2020). Application of statistical models: parameters estimation of road accident in Bangladesh*. SN Computer Science*, 1(5), 261. DOI:10.1007/s42979-020-00284-y

Shaik, Md, Islam, Md & Hossain, Q. (2021). A review on neural network techniques for the prediction of road traffic accident severity. *Asian Transport Studies*, 7(2), 100040. DOI: 10.1016/j.eastsj.2021.100040.

Sinitsyn, I.S., Sulitsky, M.V., Parygin, D.S., Dzhagaev, V.A. & Seryakova, V.N. (2022). Using neural networks to determine the traffic situation. *Systems Analysis in Science and Education,* 2, 17–22.

Song, X., Liu, Y., Xue, L., Wang, J., Zhang, J., Wang, J., Jiang, L. & Cheng, Z. (2020). Time-series well performance prediction based on long short-term memory (LSTM) neural network model. *Journal of Petroleum Science and Engineering*, 186, 106682. DOI: 10.1016/j.petrol.2019.106682

Tai, W., Wang, H., Chiang, C., Chien, C., Lai, K. & Huang, T. (2018). RTAIS: Road Traffic Accident Information System. *IEEE 4th International Conference on Data Science and Systems,* 1393-1397. DOI**:** 10.1109/HPCC/SmartCity/DSS.2018.00230

Tang, J., Zheng, L., Han, C., Yin, W., Zhang, Y., Zou, Y. & Huang, H. (2020). Statistical and machine-learning methods for clearance time prediction of road incidents: a methodology review. *Analytic Methods in Accident Research*, 27, 100123. DOI:10.1016/j.amar.2020.100123

Vanitha, R. & Swedha, M. (2023). Prediction of Road Accidents using Machine Learning Algorithms. *Middle East Journal of Applied Science & Technology (MEJAST),* 6(2), 64-75, DOI: 10.46431/MEJAST.2023.6208

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. & Polosukhin, I. (2017). Attention is All you Need. *31st Conference on Neural Information Processing Systems (NIPS). Advances in Neural Information Processing Systems*, 30. Curran Associates, Inc. arXiv:1706.03762.

Wadhahi, N., Hussain, S., Yosof, K., Hussain, S. & Singh, A. (2018). Accidents Detection and Prevention System to reduce Traffic Hazards using IR Sensors. *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)* (ICRITO), 737-741. DOI: 10.1109/ICRITO.2018.8748458.

Wei, C. H. & Lee, Y. (2007). Sequential Forecast of Incident Duration Using ANN Models. *Accident Analysis and Prevention*, 39(5), 944-954. DOI: 10.1016/j.aap.2006.12.017

Wenqi, L., Luo, D., & Yan, M. (2017). A model of traffic accident prediction based on convolutional neural network. *2017 2nd IEEE International Conference on Intelligent Transportation Engineering (ICITE)*, 198-202. DOI:10.1109/ICITE.2017.8056908

WHO report "Speed Management" (2017). Retrieved from: www.who.int/mediacentre/news/releases/2017/speed-management-cities/ru/.

Xie, T., Shang, Q. & Yu, Y. (2022). Automated Traffic Incident Detection: Coping With Imbalanced and Small Datasets. *IEEE Access,* 10(1287), 1-1. DOI: 10.1109/ACCESS.2022.3161835

Yuan, J., Abdel-Aty, M., Gong, Y. & Cai, Q. (2019). Real-time crash risk prediction using Long short-term memory recurrent neural network. *Transportation Research Record Journal of the Transportation Research Board,* 2673(1), 036119811984061. DOI: 10.1177/0361198119840611

Zhang, J., Wang, J. & Fang, S. (2019). Prediction of urban expressway total traffic accident duration based on multiple linear regression and artificial neural network. *5th International Conference on Transportation Information and Safety (ICTIS),* pp. 503-510. DOI**:** 10.1109/ICTIS.2019.8883690

Zhang, L., Liu, Q., Yang, W., Wei, N. & Dong, D. (2013). An improved K-nearest neighbor model for short-term traffic flow prediction. *Procedia - Social and Behavioral Sciences*, 96, 653-662. DOI: 10.1016/j.sbspro.2013.08.076

Zheng, M., Li, T., Zhu, R., Chen, J., Ma, Z., Tang, M., Cui, Z. & Wang, Z. (2019). Traffic Accident's Severity Prediction: A Deep-Learning Approach-Based CNN Network. *IEEE Access, 7,* 39897-39910. DOI: 10.1109/ACCESS.2019.2903319.

Zhu, L., Guo, F., Krishnan, R., & Polak, J.W. (2018). A Deep Learning Approach for Traffic Incident Detection in Urban Networks. *2018 21st International Conference on Intelligent Transportation Systems (ITSC),* 1011-1016. DOI**:** 10.1109/ITSC.2018.8569402

Zhu, W., Wu, J., Fu, T., Wang, J., Zhang, J. & Shangguan, Q. (2021). Dynamic prediction of traffic incident duration on urban expressways: a deep learning approach based on LSTM and MLP. *Journal of Intelligent and Connected Vehicles*, 4(2), 80-91. https://doi.org/10.1108/JICV-03-2021-0004

Zikratova, T. V. & Zikratov, I. A. (2020). Application of a neural network for detecting emergency situations on the roads. *Scientific and technical bulletin of information technologies, mechanics and optics,* 20(2), 301-305. DOI: 10.17586/2226-1494-2020-20-2-301-305