Overview of Publicly-Available Data Sources on Road Traffic Accidents in Russia

Alexey Girin[1], Nikolay Teslya[2][a] and Nikolay Shilov[2][b]

[1]*ITMO University, St. Petersburg, Russia*

[2]*SPC RAS, St. Petersburg, Russia*

Keywords: Road Traffic Accident, Risk Factors, Data Sources, Data Analysis.

Abstract: It is very important to develop measures to prevent traffic accidents to increase road safety. The paper provides identification and description of various publicly-available sources of data related to road traffic accidents in Russia. All the described data could be combined to support an detailed analysis of accidents. A review and classification of risk factors associated with road accidents was conducted, and the data required for accident analysis as well as publicly accessible sources of such data in Russia were described. Additionally, a review of the methods used for analyzing and predicting accidents was undertaken.

1 INTRODUCTION

Road traffic accidents (RTAs) represent a significant socio-economic problem that is relevant both globally and in Russia. According to official data, in 2023 in Russia there were more than 132.4 thousand RTA (4.5% more than a year earlier), in these accidents more than 14.5 thousand people died (2.3% more than a year earlier) and more than 166.5 thousand were injured (4.3% more than a year earlier) (Main Directorate for Traffic Safety of the Ministry of Internal Affairs of Russia, 2024). RTAs also lead to material losses, encompassing both direct costs associated with property damage and medical care, and indirect costs related to workforce loss and reduced labour productivity (Másilková, 2017).

To ensure road safety, it is essential to develop measures to prevent RTAs. One of the main aspects of road safety involves analyzing RTAs. The knowledge gained about the causes of RTAs through this analysis can increase the efficiency of the decision-making process related to road safety (Goniewicz et al., 2016). One of the important steps in RTA analysis is the collection and preparation of data for analysis. This involves describing data related to accident-prone factors, and identifying sources of such data. The accuracy of analysis results directly depends on the ability to integrate data from diverse sources, as well as the completeness and reliability of those sources (Gutierrez-Osorio and Pedraza, 2020).

Despite the fact that some researchers manage to cooperate with public authorities in Russia in order to gain access to data on RTA (Kudryavtsev et al., 2013), the general tendency is that most of the RTA data are used only by public authorities for internal analysis and are not publicly available. This lack of an accessible tool for researchers to access comprehensive, open data on RTAs in the country negatively impacts the quality of RTA analysis results. The aim of this paper is to identify and describe various data sources on RTAs in Russia that are open for public access and can be combined for further analysis by researchers.

The paper is organized as follows. Section 2 outlines the methods used in this study. Section 3 reviews the RTA risk factors. Section 4 describes the RTA data sources specifically in Russia. In section 5, methods of RTA analysis and prediction are reviewed. A discussion of the study follows in section 6, and the paper concludes in section 7.

2 RESEARCH METHODS

Firstly, we systematizes the potential causes of RTAs described in the review literature of RTA risk factors, and identify data associated with these factors. Based on this review, we identify a classification of RTA risk factors into three categories: human factors, environ-

[a] https://orcid.org/0000-0003-0619-8620
[b] https://orcid.org/0000-0002-9264-9127

Girin, A., Teslya, N. and Shilov, N.
Overview of Publicly-Available Data Sources on Road Traffic Accidents in Russia.
DOI: 10.5220/0012737100003702

In *Proceedings of the 10th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2024)*, pages 480-487
ISBN: 978-989-758-703-0; ISSN: 2184-495X

mental factors, and vehicle factors. Next, we review papers on RSA research to identify data sources that correlate to the data usually used in RTA analysis. These sources are categorized as follows: public authorities, onboard devices, road infrastructure, and social media. Using this categorization, we search and analyze publicly available data sources on RTAs in Russia. To identify RTA data sources from public authorities in Russia, we examine regulations regarding RTA data registration and information resources provided by those authorities. We also review mapping and weather services that provide additional data for contextualising RTAs. Finally, we review the methods that use these data to analyse and predict RTAs.

3 OVERVIEW OF RTA RISK FACTORS

Road traffic, as a complex dynamic system, includes several interrelated elements like traffic participants (drivers, passengers, pedestrians), vehicles and the external environment, which consists of the road network, weather and time conditions, as well as the surrounding terrain and traffic flow (Fig. 1) (Vogel and Bester, 2005). Disturbances in the functioning of these elements can lead to RTAs. Each traffic element defines a group of RTA risk factors.

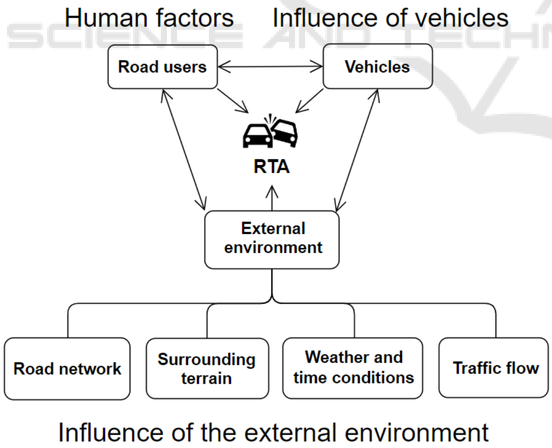

Figure 1: RTA risk factors.

3.1 Human Factors

Human-related RTA risk factors are divided according to two criteria: the intentionality of the road user's behaviour and the duration of the factor's influence (Fig. 2) (Bucsuházy et al., 2020).

Long-term RTA risk factors caused by intentional actions are associated with negative habits and behavioural characteristics of road users. Such factors include:

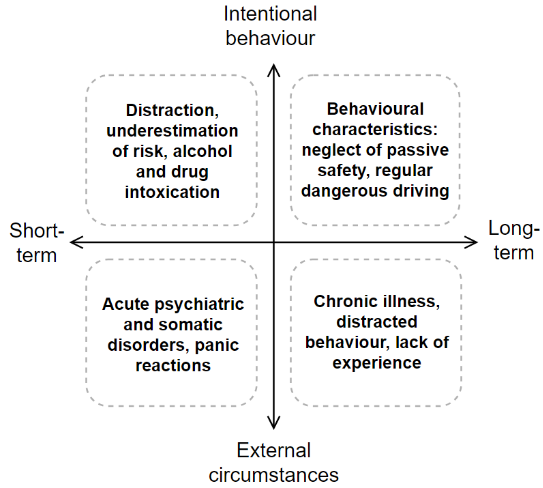

Figure 2: RTA. Human factors.

- Neglect of passive safety (ignoring the use of child seats and seat-belts) (Fattakhov, 2018).
- Regular reckless or dangerous behaviour (speeding and failure of drivers to maintain lateral spacing) (Alonso Plá et al., 2013).

The actions of road users that are consciously taken immediately before and cause a accident are the short-term accident risk factors caused by the intentional behaviour of road users:

- Underestimation of risk (e.g. overtaking on an unsafe road section, overtaking in heavy traffic, crossing the road in an inappropriate place) (Parker et al., 1995).
- Distraction while driving (e.g. using a mobile phone while driving) (Horsman and Conniss, 2015).
- Alcohol and drug intoxication (Rodionova et al., 2022).

Persistent conditions that adversely affect road users identify the following long-term risk factors for RTA: medical causes (chronic neurological diseases, narcolepsy) (Lindsay and Baldock, 2008), distracted behaviour (Parker et al., 1995), insufficient driving experience (Hu et al., 2020).

Short-term factors caused by unintentional driving behaviour include: mental and somatic abnormalities (acute stress, bouts of illness) (Taylor and Dorn, 2006), fatigue state (microsleep) (De Mello et al., 2013), panic reaction (Bucsuházy et al., 2020).

The following data on RTA's participants are correlated with the identified risk factors: common information (age, driving experience), medical information (chronic diseases, mental and somatic abnormalities), condition and behaviour of the participants.

3.2 Influence of the External Environment

Traffic is influenced by the external environment, such as the condition of the road network, weather and time conditions, the characteristics of the surrounding area, and traffic flow.

RTA risk factors caused by the condition of the road network include (Fattakhov, 2018; Wang et al., 2013): the arrangement of roads (location of traffic lights and regulatory signs, etc.), road surface conditions (low level of pavement, presence of potholes), road geometries, conditions of artificial lighting, and the presence of foreign objects in traffic areas.

RTA risk factors caused by weather and time conditions (Zou et al., 2021; Hazaymeh et al., 2022): precipitation, wind, fog; time of day, and natural lighting conditions.

RTA risk factors caused by the characteristics of the surrounding terrain (Hazaymeh et al., 2022): terrain relief (flat, hilly, or mountainous), and the density and type of buildings and other constructions.

RTA risk factors caused by the characteristics of traffic flow: speed regime and traffic intensity are separately identified (Zhang et al., 2020).

Consequently, the following data on the environmental conditions at the accident site during the RTA correlate with the identified risk factors: the state and layout of the road infrastructure, the time and weather, terrain, the presence, type, and density of points of interest, and the characteristics of the traffic flow.

3.3 Influence of Vehicles

RTA risk factors related to vehicles (Zovak et al., 2016): wear and tear of the vehicle (correlated with the following data: vehicle mileage, period of use); unsatisfactory technical condition of the vehicle (data on the presence of damage and faults of the vehicle).

4 OVERVIEW OF RTA DATA SOURCES

RTA data sources that are used in accident analysis and prediction studies:

- Public Authorities: generally collect data on the type, location, time and participants, as well as data on the causes and consequences of the accident (Rabbani et al., 2022).
- On-board Devices: devices that are mounted on a vehicle to collect vehicle characteristics and driver state information (e.g. drowsiness, anxiety, distraction) (Gutierrez-Osorio and Pedraza, 2020; Chand et al., 2021).
- Road Infrastructure Devices: technical devices that monitor the movement of vehicles: radars, cameras, loop traffic detectors. (Gutierrez-Osorio and Pedraza, 2020).
- Social Media: information extracted from accident publications has been used to identify RTA locations and to identify clusters of high accident rates (Chand et al., 2021).

4.1 RTA Data Sources in Russia

The main source of data on RTA in Russia are organisations that keep records of RTA: internal affairs authorities; road owners; organisations owning vehicles; and medical organisations. Table 1 provides information about the data on RTA collected by these organisations. The data on RTA records from internal affairs authorities are partially open for public access. These data can be accessed through a specialised service ”Indicators of the Road Safety State” provided by the Traffic Police of the Russian Ministry of Internal Affairs and described in the subsection 4.1.1. This service is the primary, and in most cases, the only source of data on traffic accidents used in the analysis of RTAs in Russia.

For instance, the research (Donchenko et al., 2020) considered the potential for creating a model that could predict traffic accidents based on data about accidents in Russia between the beginning of 2015 and April 2018. These data was provided by this service. Another example of using this service is research (Kasatkina and Vavilova, 2023), which developed an intelligent system that determines the concentration of road accidents in the Udmurt Republic (Russia) through the use of accident data from 2022.

Some accident analysis studies conducted in Russia use accident data acquired from the ”Indicators of the State of Road Safety” service in conjunction with information from various other sources. For example, the authors of the study (Fatkulin et al., 2017) used data obtained from the service and data extracted from posts in VKontakte social network during 2017 when developing a system for monitoring accidents through real-time analysis of social media.

For a more in-depth contextualisation of RTA, it is necessary to consult additional data sources. These include road and weather conditions at the time of the RTA, the condition and behaviour of the participants in the RTA, and the environment of the RTA scene.

Table 1: RTA data in Russian organisations.

Organisations	Description of RTA record data	Public access
Internal affairs authorities	General information about the RTA, information about the external environment (weather conditions, road network state, surroundings of the accident site), and participants and vehicles involved in the RTA	Partial
Road owners	Data on the condition of the road network in the area of the RTA	No
Organisations owning vehicles	General information about the accident and results of internal investigation: information about driver and injured (severity), information about vehicles, details of accident (violation of traffic rules fact), consequences of accident (material losses, number of deaths and injuries)	No
Medical organisations	Data on those killed and injured in RTA who were treated by a medical organisation	No

These sources are described in the following subsections.

4.1.1 Indicators of the State of Road Safety

"Indicators of the State of Road Safety" is a service provided by the Russian Ministry of Internal Affairs Traffic Police. This service provides statistical data on the state of accident rates and detailed accident cards (Main Directorate for Traffic Safety of the Ministry of Internal Affairs of Russia, 2024). The cards contain detailed information on the circumstances of specific RTA. The cards follow a fixed structure with sections for general information, scheme of the accident, road conditions on the scene, data on vehicles, and other participants (Table 2). The service does not provide information on permissible values for each field of the card. However, this information can be obtained from regulations and recommendations which are publicly available. The procedure for filling out the fields is described in the "Recommendations for recording and analysing RTA on Russian highways" from the Federal Road Agency Rosavtodor. Table 2 gives a description of each section of the cards. Maps can be downloaded separately for each region of the Russian Federation. The period of time for which all cards can be downloaded at once is limited to one month. The service only allows downloading cards for those RTAs that occurred before 2015. The following document formats are supported: PDF, XLS, CSV, and XML. The service is a major source of data on RTAs in Russia, providing the most comprehensive basic information about the circumstances of these RTAs.

Table 2: Structure of RTA card.

Section Title	Section Description
General information	Place and time of the RTA, number of participants and number of vehicles
Road conditions	Description of objects at the accident site and near it, road conditions and deficiencies, weather conditions, and other information about the surroundings of the accident
Vehicles	Description of vehicles involved in the accident, presence or absence of technical faults
RTA participants	Categories of participants (driver, passenger, cyclist, etc.), level of injury, breach of road traffic rules
RTA diagram	A graphical representation of the circumstances surrounding the accident (the location and direction of motion of participants and vehicles during the accident)

4.1.2 Road Fund Control System

Road Fund Control System from ROSDORNII provides information about roads and road papers plans, as well as data on RTA (The Federal Autonomous Institution "Russian Road Scientific-Research Institute", 2024). The data can be accessed through the user interface; the system does not provide a tool for uploading them. The system provides detailed up-to-date data on both federal and regional roads and local roads. Table 3 provides a description of the information page sections for road data, some of the fields (road characteristics) are tempered, i.e. their start and

Table 3: Structure of the road information page in Road Fund Control System.

Section Title	Section Description
Legal information	Road owner and operating organisation
Project activity	Information on the types of papers on a particular road
Technical data	Road class and category, pavement type, number of lanes, maximum speed, climatic zone, axle load, capacity, carriageway width, subgrade width
Normative road condition	Compliance of the road with the normative condition
Safety	Availability of the accident-hazardous status, accidents with dead and injured people on the road

end times are specified for the values of these fields. The data on RTA provided by the system correspond in structure and content to the data of the "Indicators of the State of Road Safety" service. Data on road plan documents for a particular motorway include: type of documents, timing of documents, technical indicators of road sections before and after documents, etc. The information system Road Fund Control System can be used as an additional source of data on RTA, specifically to obtain information about the condition of the road network in the area where the accident occurred.

4.1.3 Mapping Services

The OpenStreetMap (OSM) service (OpenStreetMap Contributors, 2024a), which is the leading open-source of geographic information (Mooney et al., 2017), can be used to collect additional information about the accident site environment. The process of defining the accident site environment using the OSM can be described as follows:

1. Define a set of objects types that may affect traffic safety.
2. Match each type of objects from the obtained set with a template for search in OSM.
3. Implement a function that searches for objects of specific types within the given radius around the coordinates of the accident.

To obtain OSM data it is possible to download a snapshot for a certain region using Geofabrik tool (Geofabrik GmbH Karlsruhe, 2024). An alternative option is to use the Overpass service (OpenStreetMap Contributors, 2024b), which provides an API for obtaining OSM objects by user requests. The obtained data can be used in addition to the data from the "road conditions" section of RTA cards.

4.1.4 Sources of Weather Data

Data on weather conditions at the location and time of the accident can be obtained directly from the accident card. For a detailed report on weather conditions it may be necessary to use external services. The main criteria for choosing such a service are: the ability to access historical weather data based on geographic coordinates and time, the availability of data for Russia, the maximum data update interval should not exceed one hour (to ensure the most accurate readings), the ability to retrieve values for a specific range of weather variables. One of the services that meets these criteria is Open-Meteo, which provides an API for accessing historical weather data up to 10,000 times per day for personal use, making it the best option compared to similar services (Open-Meteo contributors, 2024).

4.1.5 Court Websites

Data on the circumstances of a particular RTA can be obtained from the text of a court judgment on the accident. The sources of documents with verdicts in RTA cases include court websites, where the texts of court decisions are published. When working with these sources, the following difficulties arise: the need to compare court decisions with specific RTAs; the need to handle unstructured texts from court decisions; and the dispersion of places where court decisions are posted. With a court decision text analysis it is possible to determine information about the participants in an accident (a description of their behaviour during the accident, their medical condition, the fact of alcohol or drug intoxication, and any injuries received as a result of the RTA), information about vehicles (model, period of use, and presence and type of mechanical malfunctions), and the state of the surrounding environment (the weather and time conditions and condition of the road network).

4.1.6 Other Sources

Alternative sources of open data on RTAs in Russia include publications on social media and in the public media, as well as publicly available recordings from onboard video recorders. Further research is required to explore the possibilities for searching, interpreting and comparing these data with specific RTA.

5 METHODS OF RTA ANALYSIS AND PREDICTION

In order to further use of RTA data from various public sources for RTA analysis and prevention purposes, it is necessary to have an understanding of the current methods for analyzing and predicting RTAs. It is worth noting that a significant number of techniques used in analyzing and predicting RTAs are related to artificial intelligence and machine learning.

5.1 Identification of High Accident Clusters

One of the steps in road safety is to identify high-accident clusters in a given area. This task can be formulated as a search for road segments with the highest accident density. To solve this problem, clustering methods based on the analysis of historical data on RTA are used (Chang et al., 2022). For example, in the study (Moosavi et al., 2019) DBSCAN clustering algorithm is used to identify clusters of higher accident density. Another common method is Kernel Density Estimation (KDE), which allows to calculate the density of RTA points in the study area and identify areas with the highest density values (Santos et al., 2021).

5.2 Identification of RTA Risk Factors

The high fatality rate in RTA is attributed to the lack of awareness of RTA risk factors, and many studies have aimed at identifying these factors, which is also an important step towards road safety (Alkheder et al., 2020). For example, the analysis of RTA data in (Wang et al., 2021) is helped to identify RTA risk factors and to propose a model to quantify them in order to compare the probability of RTA in different scenarios. Part of the researches in this area focuses on evaluating the impact of RTA risk factors on the severity of the accident consequences, which can be expressed through the number of fatalities and casualties in the accident, as well as the extent of injury to victims. For example, in (Eboli et al., 2020) this problem is solved using Binary Logistic Regression, and in (Alkheder et al., 2020) the decision tree, Support Vector Machine (SVM) and Bayesian network are used to estimate the degree of influence of risk factors on accident severity.

5.3 RTA Prediction

Challenges in the field of RTA prediction include predicting the occurrence and estimating the severity of RTA. One of the advanced trends in this area is the application of deep learning algorithms (Gutierrez-Osorio and Pedraza, 2020). For example, (Rolison, 2020) presented a method for predicting the severity of RTA based on encoding a matrix of RTA data into special grey images, which are then used as inputs for a convolutional neural network. The results of the study showed that the proposed method of predicting RTA severity using a convolutional neural network provides higher accuracy compared to other methods such as K-nearest neighbours algorithm, SVM and recurrent neural networks.

Rough set theory can be used in RTA data analysis to assess the influence of risk factors on RTA outcomes. It can also be used to reduce the number of attributes associated with risk factors and to detect hidden relationships between the data and for classification tasks (Jianfeng et al., 2019).

In papers on RTA analysis and prediction, there is a tendency to use multiple data sources to improve the accuracy of the results. Researchers explicitly point out that the success of prediction models depends mainly on how data from different sources can be integrated (Chand et al., 2021; Suat-Rojas et al., 2022). In addition to the above studies, another example of such integration is the paper of (Marcillo et al., 2022), which combined heterogeneous data on traffic accidents, weather conditions and hotspots. Based on a deep neural network, a real-time traffic accident prediction model was built, which performs better than other models.

6 DISCUSSION

6.1 RTA Risk Factors

The following RTA factors were identified: human factors; external environment; vehicle factors. Data that relate to these factors can be used to provide a comprehensive understanding of the accident context.

6.2 RTA Data Sources

The sources of RTA data include: public authorities, onboard devices, road infrastructure devices, and social media. The main source of RTA data in Russia is the traffic police service (public authorities), which is open to the public and provides basic information about the circumstances of the RTAs, accident participants, and vehicles. Other data collected by public authorities on RTAs are used only for internal purposes and are closed to the public. Additional sources of public RTA data that could be integrated into a unified system are:

- Road Fund Control System: road network data. Can be used to determine the road condition at the location of the accident, as well as to assess the intensity of traffic flow.
- OSM mapping service: can be used to obtain data on the environment of the accident site.
- Weather APIs: can be used to determine weather conditions at the location and time of the accident.
- Court websites and aggregators of court decisions can be used to obtain information on the circumstances of a road accident, a detailed description of the state and behaviour of RTA participants, and characteristics of vehicles and the external environment. The data is unstructured and requires pre-processing.
- Publications on social media, as well as publicly available recordings from onboard devices (video recorders), need further research to explore the possibilities of searching and interpreting data. These sources are not included in this paper.

6.3 Methods of RTA Analysis and Prediction

The following road safety tasks were identified: identification of clusters of increased accident rate (DBSCAN clustering algorithm and KDE); identification of RTA risk factors (decision tree, SVM and Bayesian network); accident prediction (convolutional neural network and recurrent neural network).

7 CONCLUSION

This paper classifies risk factors for RTAs and describes data that can be used in RTA analysis. The sources of RTA data were reviewed, and the publicly available sources in Russia were searched and described. Additionally, a review was conducted on methods for analysing and predicting RTAs. In the future work, it is planned to implement a software system to automatically collect RTA data from the public sources, develop methods of RTA analysis and prediction based on the data obtained.

ACKNOWLEDGEMENTS

The study was carried out within State Research, project number FFZF-2022-0005.

REFERENCES

Alkheder, S., AlRukaibi, F., and Aiash, A. (2020). Risk analysis of traffic accidents' severities: An application of three data mining models. *ISA transactions*, 106:213–220.

Alonso Plá, F. M., Esteban Martínez, C., Calatayud Miñana, C., and Sanmartín, J. (2013). Speed and road accidents: behaviors, motives, and assessment of the effectiveness of penalties for speeding. *American Journal of Applied Psychology, 2013, vol. 1, num. 3, p. 58-64*.

Bucsuházy, K., Matuchová, E., Zůvala, R., Moravcová, P., Kostíková, M., and Mikulec, R. (2020). Human factors contributing to the road traffic accident occurrence. *Transportation research procedia*, 45:555–561.

Chand, A., Jayesh, S., and Bhasi, A. (2021). Road traffic accidents: An overview of data sources, analysis techniques and contributing factors. *Materials Today: Proceedings*, 47:5135–5141.

Chang, H., Li, L., Huang, J., Zhang, Q., and Chin, K.-S. (2022). Tracking traffic congestion and accidents using social media data: A case study of shanghai. *Accident Analysis & Prevention*, 169:106618.

De Mello, M. T., Narciso, F. V., Tufik, S., Paiva, T., Spence, D. W., BaHammam, A. S., Verster, J. C., and Pandi-Perumal, S. R. (2013). Sleep disorders as a cause of motor vehicle collisions. *International journal of preventive medicine*, 4(3):246.

Donchenko, D., Sadovnikova, N., and Parygin, D. (2020). Prediction of road accidents' severity on russian roads using machine learning techniques. In *Proceedings of the 5th International Conference on Industrial Engineering (ICIE 2019) Volume II 5*, pages 1493–1501. Springer.

Eboli, L., Forciniti, C., and Mazzulla, G. (2020). Factors influencing accident severity: an analysis by road accident type. *Transportation research procedia*, 47:449–456.

Fatkulin, T., Butakov, N., Krikunov, A., and Voloshin, D. (2017). Accident monitoring framework based on online social network sensing. *Procedia computer science*, 119:278–287.

Fattakhov, T. (2018). Traffic injuries in russia and its factors.

Geofabrik GmbH Karlsruhe (2024). Geofabrik, official website. retrieved february 22, 2024. https://www.geofabrik.de/.

Goniewicz, K., Goniewicz, M., Pawłowski, W., and Fiedor, P. (2016). Road accident rates: strategies and programmes for improving road traffic safety. *European Journal of Trauma and Emergency Surgery*, 42(4):433–438.

Gutierrez-Osorio, C. and 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.

Hazaymeh, K., Almagbile, A., and Alomari, A. H. (2022). Spatiotemporal analysis of traffic accidents hotspots based on geospatial techniques. *ISPRS International Journal of Geo-Information*, 11(4):260.

Horsman, G. and Conniss, L. R. (2015). Investigating evidence of mobile phone usage by drivers in road traffic accidents. *Digital Investigation*, 12:S30–S37.

Hu, L., Bao, X., Wu, H., and Wu, W. (2020). A study on correlation of traffic accident tendency with driver characters using in-depth traffic accident data. *Journal of advanced transportation*, 2020:1–7.

Jianfeng, X., Hongyu, G., Jian, T., Liu, L., and Haizhu, L. (2019). A classification and recognition model for the severity of road traffic accident. *Advances in Mechanical Engineering*, 11(5):1687814019851893.

Kasatkina, E. and Vavilova, D. (2023). Identification and mitigation of road traffic accident concentration areas by using cluster analysis. In *2023 5th International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA)*, pages 247–251. IEEE.

Kudryavtsev, A. V., Kleshchinov, N., Ermolina, M., Lund, J., Grjibovski, A. M., Nilssen, O., and Ytterstad, B. (2013). Road traffic fatalities in arkhangelsk, russia in 2005–2010: Reliability of police and healthcare data. *Accident Analysis & Prevention*, 53:46–54.

Lindsay, V. and Baldock, M. (2008). Medical conditions as a contributing factor in crash causation.

Main Directorate for Traffic Safety of the Ministry of Internal Affairs of Russia (2024). Information on road safety indicators. retrieved february 22, 2024. http://stat.gibdd.ru/.

Marcillo, P., Valdivieso Caraguay, Á. L., and Hernández-Álvarez, M. (2022). A systematic literature review of learning-based traffic accident prediction models based on heterogeneous sources. *Applied Sciences*, 12(9):4529.

Mooney, P., Minghini, M., et al. (2017). A review of openstreetmap data. *Mapping and the citizen sensor*, pages 37–59.

Moosavi, S., Samavatian, M. H., Parthasarathy, S., Teodorescu, R., and Ramnath, R. (2019). Accident risk prediction based on heterogeneous sparse data: New dataset and insights. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 33–42.

Másilková, M. (2017). Health and social consequences of road traffic accidents. *Kontakt*, 19(1):e43–e47.

Open-Meteo contributors (2024). Open-meteo, official website. retrieved february 22, 2024. https://open-meteo.com/en/docs/historical-weather-api.

OpenStreetMap Contributors (2024a). Openstreetmap. retrieved february 22, 2024. https://www.openstreetmap.org.

OpenStreetMap Contributors (2024b). Overpass api. retrieved february 22, 2024. https://wiki.openstreetmap.org/wiki/Overpass_API.

Parker, D., Reason, J. T., Manstead, A. S., and Stradling, S. G. (1995). Driving errors, driving violations and accident involvement. *Ergonomics*, 38(5):1036–1048.

Rabbani, M. B. A., Musarat, M. A., Alaloul, W. S., Ayub, S., Bukhari, H., and Altaf, M. (2022). Road accident data collection systems in developing and developed countries: a review. *International Journal of Integrated Engineering*, 14(1):336–352.

Rodionova, M., Skhvediani, A., and Kudryavtseva, T. (2022). Prediction of crash severity as a way of road safety improvement: The case of saint petersburg, russia. *Sustainability*, 14(16):9840.

Rolison, J. J. (2020). Identifying the causes of road traffic collisions: Using police officers' expertise to improve the reporting of contributory factors data. *Accident Analysis & Prevention*, 135:105390.

Santos, D., Saias, J., Quaresma, P., and Nogueira, V. B. (2021). Machine learning approaches to traffic accident analysis and hotspot prediction. *Computers*, 10(12):157.

Suat-Rojas, N., Gutierrez-Osorio, C., and Pedraza, C. (2022). Extraction and analysis of social networks data to detect traffic accidents. *Information*, 13(1):26.

Taylor, A. H. and Dorn, L. (2006). Stress, fatigue, health, and risk of road traffic accidents among professional drivers: the contribution of physical inactivity. *Annu. Rev. Public Health*, 27:371–391.

The Federal Autonomous Institution "Russian Road Scientific-Research Institute" (2024). Road fund control system. retrieved february 22, 2024. https://xn--d1aluo.xn--p1ai/.

Vogel, L. and Bester, C. (2005). A relationship between accident types and causes. *SATC 2005*.

Wang, C., Quddus, M. A., and Ison, S. G. (2013). The effect of traffic and road characteristics on road safety: A review and future research direction. *Safety science*, 57:264–275.

Wang, X., Qu, Z., Song, X., Bai, Q., Pan, Z., and Li, H. (2021). Incorporating accident liability into crash risk analysis: A multidimensional risk source approach. *Accident Analysis & Prevention*, 153:106035.

Zhang, Y., Lu, H., and Qu, W. (2020). Geographical detection of traffic accidents spatial stratified heterogeneity and influence factors. *International journal of environmental research and public health*, 17(2):572.

Zou, Y., Zhang, Y., and Cheng, K. (2021). Exploring the impact of climate and extreme weather on fatal traffic accidents. *Sustainability*, 13(1):390.

Zovak, G., Kučinić, T., and Ševo, I. (2016). Importance of technical inspection of vehicles after traffic accidents. *Trans Motauto World*, 1(4):3–6.