USE OF ARTIFICIAL INTELLIGENCE FOR AUTOMATIC ANOMALY DETECTION SYSTEM IN IDS JMK OPERATION

Feasibility study - technical part

Basis for D.T1.2.5 30 07 2021

KORDIS JMK

Ing. Ondřej Hynčica et al.

CONTENTS

1. INTRODUCTION

The subject of this part of the feasibility study is the analysis of the use of artificial intelligence elements in the monitoring and control system of the Integrated Transport System of the South Moravian Region (hereinafter referred to as IDS JMK). The target application for the pilot verification of the possibilities of artificial intelligence tools (hereinafter referred to as AI) is the detection and localization of exceptional traffic conditions in the South Moravian Region in real time, i.e. detection of anomalies compared to defined timetables, including short and long-term closures.

Within the study document - technical part, the following areas are mainly processed:

- A brief introduction to artificial intelligence and the genesis of the project plan.
- The current state of the monitoring and control system of the IDS JMK.
- Definition of the task of anomaly detection and localization.
- Technical resources and knowledge requirements for AI tasks.
- Available solutions and concept of anomaly detection system.
- Framework estimate of the implementation schedule.

1.1. Definition of the study area

The study document - technical part includes information concerning the technical and technological part of the feasibility study using artificial intelligence elements for automated detection of anomalies in the operation of the JMK. The study does not contain an analysis of organisational and economic parameters and risks necessary for the overall evaluation of the investment plan, except for a framework estimate of the implementation schedule, which can be used together with other documents to determine the economic costs of the project.

The design of the concept of the anomaly detection system is focused on the specific conditions and requirements of the IDS JMK. The possible portability of the concept to transport systems of other regions or to completely different systems, e.g. industrial or medical, is due to the broad generality of the AI methods used and the partial unification of transport management systems, e.g. the use of data formats or the way of organising the operation of lines.

1.2. A brief insight into the function of artificial intelligence

Artificial intelligence is defined in the most general terms as the simulation of the manifestations of human intelligence with the help of machines [1]. A characteristic of AI is the ability to autonomously choose decisions that will lead to the achievement of a goal, based on previous experience and current measurements. The first modern age systems using AI elements date back to the 1950s

20th century and were proposed following the work of British mathematician Alan Turing (1912-1954). These were usually single-purpose demonstrative computer programs simulating board game players. The practical extension of AI into transportation and industrial applications was only made possible after the start of the 3rd millennium by the 2009 breakthrough when Nvidia increased the performance of computing stations by an order of magnitude by using new **graphics processing units** (GPUs) and then by the introduction of **tensor processing unit** (TPU) architectures by several different manufacturers.

A fundamental part of AI is Machine Learning (ML), a field of methods designed to extract and transfer knowledge (not information or data) from a real system to a computer model [2]. An illustrative example of the use of ML in transportation can be autonomous vehicles, whose firmware is adapted to the observed behaviour of human drivers in sub-modular systems, e.g. lane recognition, traffic lights, pedestrians and cyclists, horizontal and vertical traffic signs, driver fatigue detection, evaluation of imminent collisions, etc., among others, using machine learning.

In terms of application, AI is divided into strong and weak artificial intelligence. Currently, only weak AI is implemented in practice, representing systems that solve goal-specific tasks such as playing chess or using a personal electronic assistant on the web or on the phone. Strong AI, on the other hand, is represented by systems that solve complex tasks in a similar way to a human and does not require their additional interaction or control, and has been the subject of basic research to date. Examples include the aforementioned autonomous vehicle, for which legislation plays a crucial role alongside technology, or the extremely precise robotic execution of complex surgical procedures. These examples are (2021), irrespective of the marketing claims made by manufacturers, still at an early stage of development and in practice require predominantly human interaction.

For a basic illustration of the principle of machine learning and thus its characteristic features (parameters and conditions of use), it is appropriate to use the example of human learning. In order to build a functional application / educate a human individual, it is necessary to have computational capacity (computer / brain) and a high number of learning examples, on the basis of which the initially empty model (in the case of a machine, an artificial neural network, decision tree, etc.) is set to a working point. This stage is called the **learning stage** in the case of the machine and the human individual, and is iterative and computationally, data- and thus time-consuming (Fig. 1). The second stage is the **inference** stage, or prediction stage, in which the already learned model performs work in accordance with previously presented examples and, depending on the specific task, usually also in real time.

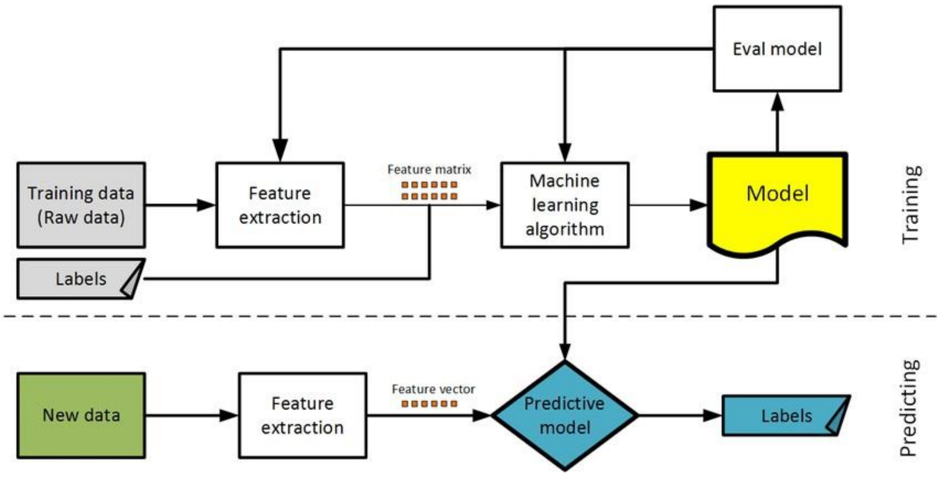


Fig. 1: Flow chart of machine learning with a teacher - learning (top) and prediction (bottom) phases.

The method of learning, simplistically and inaccurately programming, the AI model is again similar to the case of learning of a human individual either **batch** or **incremental**. Batch means that the model is set up once using an input dataset (called a dataset) and then run in the same state over and over again. In this case, the difference with traditional methods of designing systems without AI is the possibility of implementing tasks for which the analytical solution is not known. The second learning method mentioned above, i.e. incremental, allows, even after the initial batch initialization of the model, its further **adaptation** (tutoring) to new data that may differ from the original data. The incremental learning method is more complex in terms of system architecture, but has the ability to adapt to changes in the system at runtime.

Another essential characteristic of AI systems, which should be noted in the introductory chapter, especially from the practical point of view of subsequent implementation, is the so-called teacher and non-teacher learning. Learning with a teacher represents such a process of training (from the practical point of view of the end user, the term setup is more apt) a model in which for a given input data we also know its correct classification (ground-truth). For example, for a known delay ***t*** of a car in a traffic system located at time of day ***d*** and according to the geographical information obtained from a satellite navigation system (GPS/Galielo/Glonass, etc.) at a specific location ***m***, the information whether it is an anomaly or normal traffic is available already in the learning phase (the so-called dichotomous task). The second variant, i.e., learning without a teacher, is a specific way of searching for unknown structure in unannotated data, often inaccurately reported as cluster analysis.

Last but not least, it is also important to mention a general disadvantage of all AI systems, namely, for most models, low **interpretability**, i.e. the external readability of the way knowledge is stored in the model structure. As a consequence of this property, it is not possible, at least easily and quickly, to trace the reasons for a specific prediction of a model, e.g. when analysing its behaviour, because the knowledge is stored in a complex and extensive structure depending on the type of model a high number of interconnected elements. This is especially true for artificial neural networks (hereafter referred to as ANNs).

The influence and practical impact of the above mentioned basic AI features on the anomaly detection system in the IDS JMK transport system is described in the chapter Concept of the IDS JMK Anomaly Detection System.

1.3. The genesis of the use of AI in the IDS JMK monitoring system

Artificial intelligence systems are, in principle, designed for tasks for which it is impossible or too difficult to construct an exact analytical description, be it by equation, graph, table or otherwise. In particular, tasks with:

- difficult to measure internal parameters (e.g. human brain activity),
- high number of parameters, the existence of some of them may not be known at all (so-called hidden states of the system, see Markov hidden models more generally),
- extensive and usually heterogeneous structure of elements and transitions between them (e.g. modelling of social systems including transport systems).

Public transport systems, whether regional or national, always represent by definition a finite multigraph with a high number of nodes and edges, generally non-oriented, continuous and valued, with non-deterministically defined transitions between nodes (externalities).

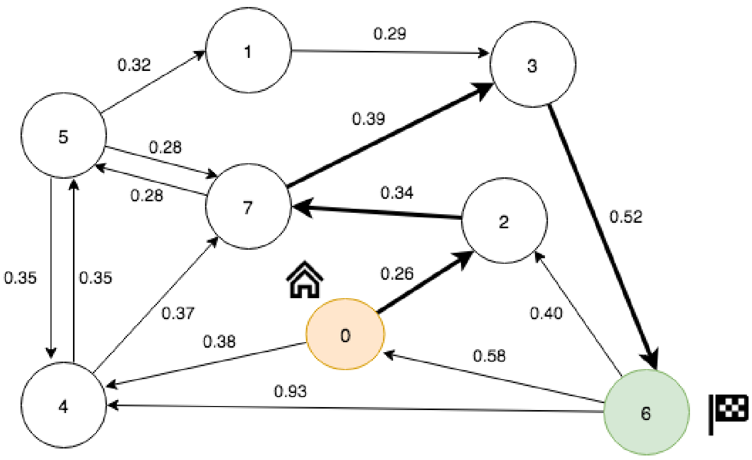


Fig. 2: Oriented, ranked, continuous and finite multigraph with seven nodes (shortest path search problem - original Dijkstra's algorithm).

In the specific case of the JMK IDS, the network of lines of the city of Brno is connected to the lines of the JMK region comprising almost 11 thousand stops (nodes of the graph), in which about 1400 entities, i.e. trains, buses, trolleybuses and trams (nodes of the graph) move in peak hours. They further enter this graph cyclically:

- planned timetable adjustments = permanent change to the internal structure of the graph,
- lockouts or mass actions, i.e. reduction/strengthening of lines = temporary change of the internal structure of the graph,
- Difficult to predict system failures, i.e. weather, accidents, congestion, outages, etc.
  = an anomaly.

From a data processing point of view, this is big data in a dynamic system, all parameters of which are not known at the time of control. **Therefore, the use of AI for anomaly detection over existing traffic system data is proposed as a potentially effective indicator of error conditions** in traffic monitoring. In general, the implementation of AI elements in control systems and data processing has an expected increasing trend since the aforementioned breakthrough in 2009. The area of AI applications in transportation infrastructure has been extensively covered in the report Artificial Intelligence Applications to Critical Transportation Issues (Washington, 2012) [4], whose internationally recognized team of authors from the Transportation Research Board identified five areas suitable for AI integration. One of the identified areas is the area of public transportation, specifically traffic flow planning and optimization, signal timing, simulation of systems with human agents, processing of observed data for consistency, and others.

The advantage of the South Moravian Region in terms of the pilot implementation of a system using AI in transport infrastructure management is the distinctive university and technologically innovative environment in which many established companies, research groups and startups are dedicated to research and use of artificial intelligence.

1.4. Sustainability in the context of AI

Given the high pace and volume of development of AI applications in almost all economic sectors, a specific agenda for sustainable AI development has been developed simultaneously by several international organisations. A general pillar is "The 2030 Agenda for Sustainable Development" [21] (2018, United Nations), focusing on the topics of regulation, transparency, ethical standards and environmental impact, or the more technical "Sustainable and Smart Mobility Strategy" [5] (2020, Brussels) of the European Commission, which is itself part of the better known European Green Deal initiative. This strategy addresses, among other things, the area of Big Data/Data Mining) and AI with machine learning as one of the flagships for mobility in the Smart-cities concept.

From the point of view of sustainability, especially economic sustainability, it can be stated with regard to a specific AI project in the IDS JMK monitoring and control system that an already designed and implemented AI system usually has lower, but at most the same requirements for operation as a system designed using traditional methods. The increased input costs for the knowledge and technical equipment required for the design of an AI system (e.g. powerful computer stations, renting a supercomputer cloud, specialised design of an AI system) are, as with all other automation systems, compensated by a combination of subsequently reduced operating costs (e.g. This is due to the increase in the level of flexibility of the system given by its intrinsic ability to adapt to changes (one-off vs. continuous AI learning, see batch and incremental AI learning in the previous subsection).

2. EXISTING TRAFFIC MONITORING AND CONTROL SYSTEM IDS JMK

2.1. Public transport in South Moravia

Public transport in the South Moravian Region is formed by connecting the regional lines of the South Moravian Region (buses and trains) with the urban public transport of the city of Brno (trams, trolleybuses, buses and boat transport) and together they form the so-called integrated transport system (IDS), at the same time with a cross-border overlap - connection of lines to Slovakia and Austria. Historically, the integration and gradual development of the South Moravian Region's transport systems took place between 2004 and 2010.

The current public transport of the IDS South Moravian Region can be described by basic numerical parameters of the regional and urban network of lines (source: Kordis JMK and [22]):

- Stops (nodes): 10826,
- Lines (edges): 322,
- Regional buses: approx. 800 vehicles at any one time (peak),
- Regional trains: about 200 trains at a time (peak),
- Brno public transport: about 400 vehicles at one moment (rush hour),
- Population served in the South Moravian Region: approx. 1.3 million from more than 700 municipalities,
- Standardised frequency of service to municipalities in JMK: 6 pairs of connections on a working day and 3 pairs on a non-working day,
- Standardized changeover quality determined by maximum changeover time: 10 min,
- Time cycle for sampling the position of vehicles in the JMK region: 6 seconds.

2.1.1. Map display of current public transport in South Moravia

The current state of public transport in the region and the city of Brno can be viewed using a dynamic online map with plotted means of transport, publicly accessible via the web interface https://mapa.idsjmk.cz. For more detailed statistics on the number of JMK IDS lines see Annex 9.3.

The dynamic map with public transport vehicles is updated every 10 seconds and can be used to interactively display selected types of vehicles and monitor their current numbers in operation (trams, trolleybuses, public transport buses, regional buses, trains and boats). Each type of vehicle has a dedicated graphic symbol (e.g. green triangles with a threedigit code = regional buses, coloured triangles with a numerical code up to 12 = trams, etc.).

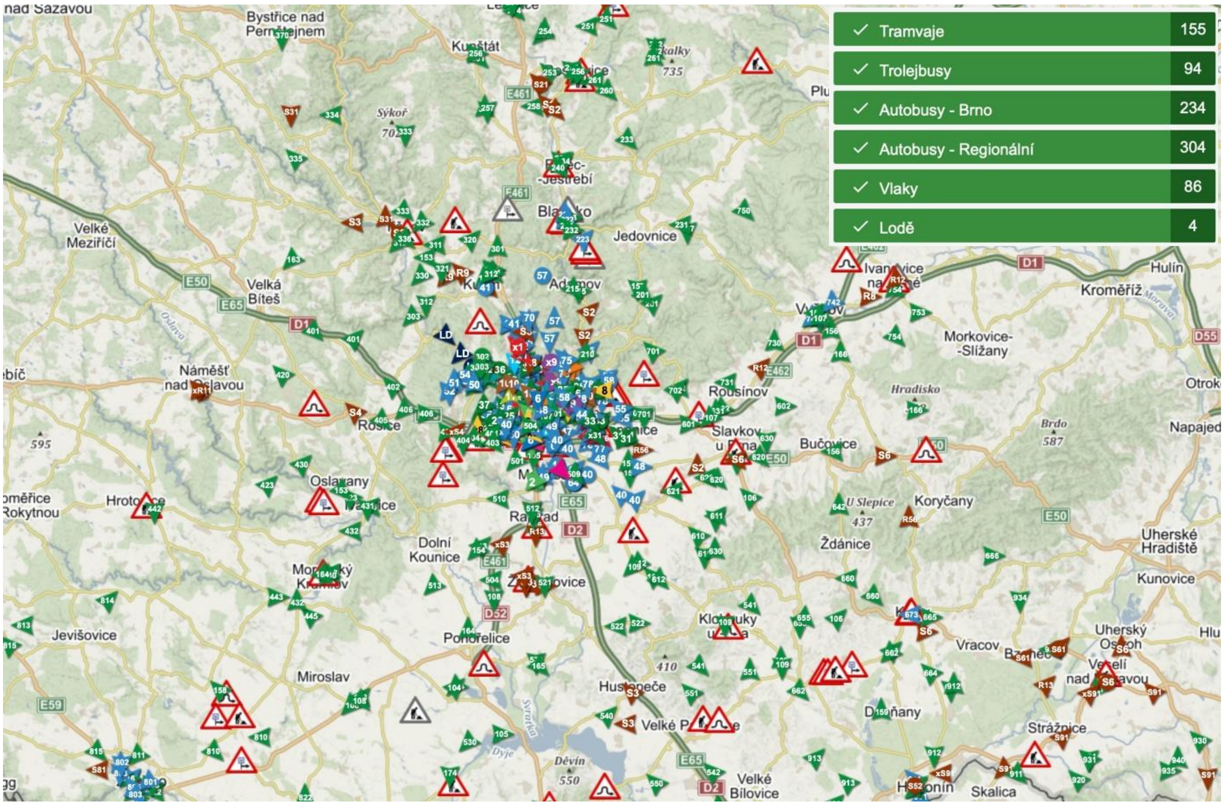


Fig. 3: Map base mapy.cz: overview of South Moravian Region - weekday traffic, approx. 900 means (06-2021).

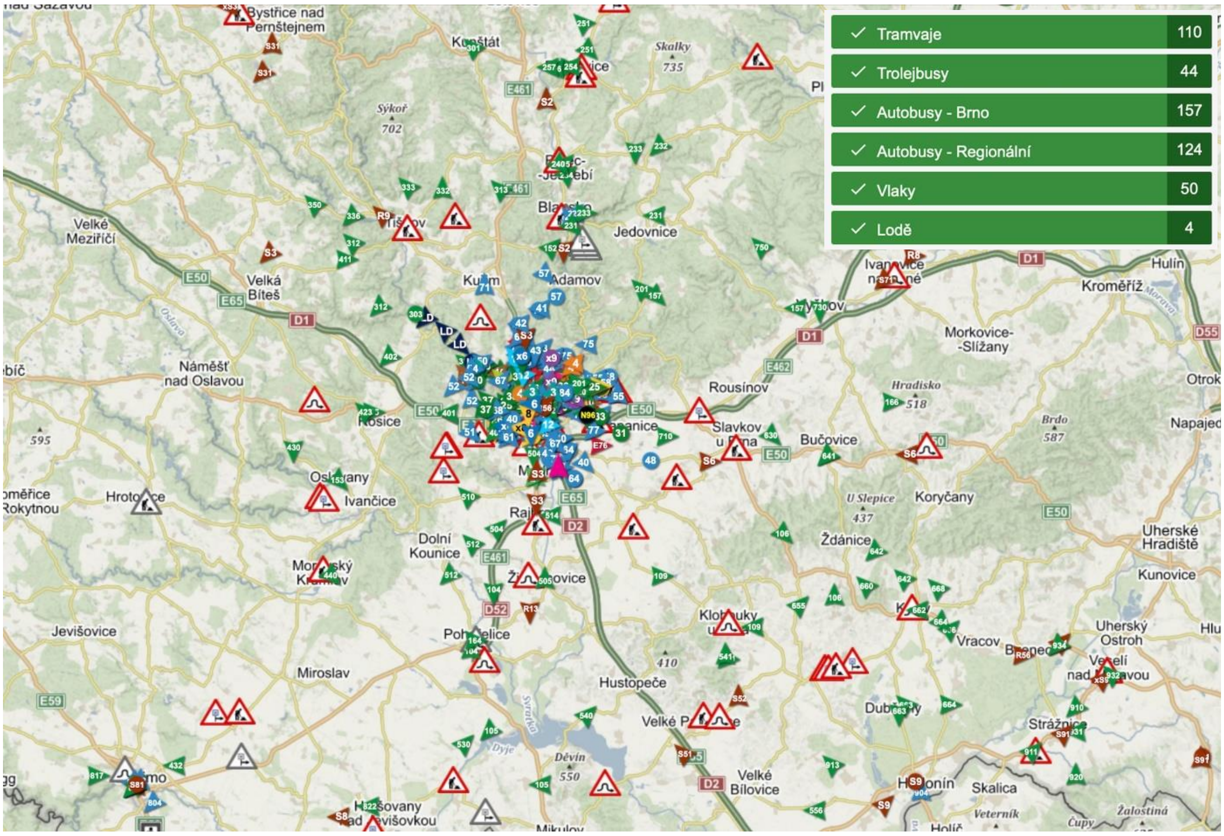


Fig. 4: Map base mapy.cz: overview of South Moravian Region - weekend traffic, approx. 500 means (06-2021).

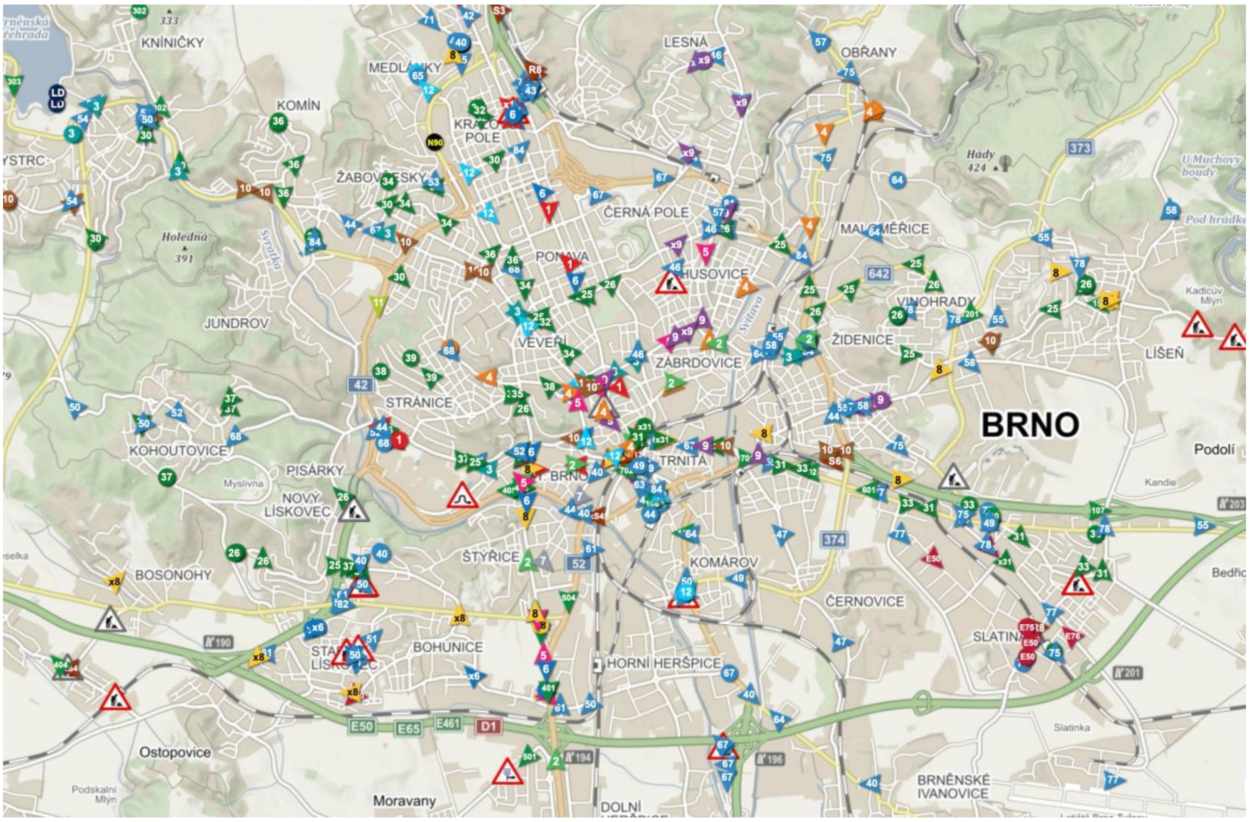


Fig. 5: Map base mapy.cz: overview of Brno public transport - weekday transport, approx. 500 means (06-2021).

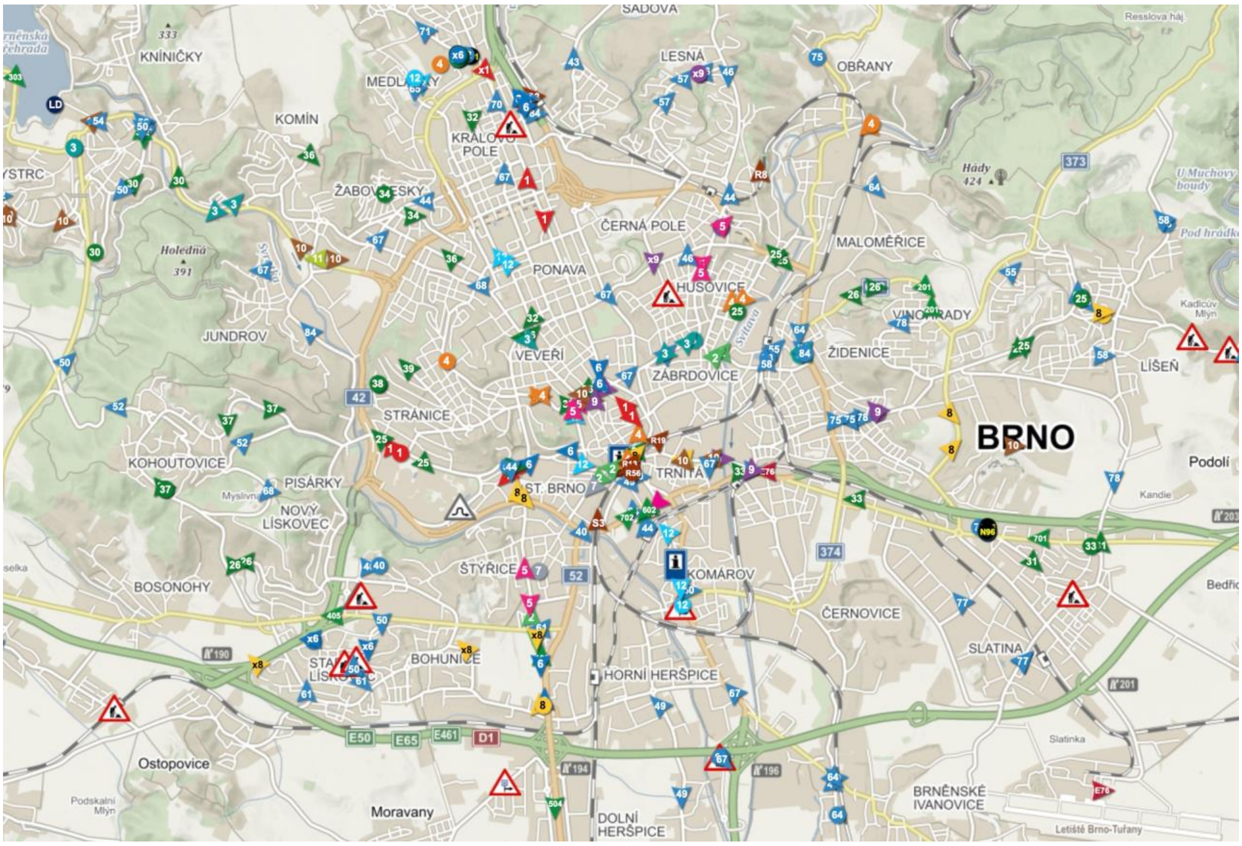


Fig. 6: Map base mapy.cz: overview of Brno public transport - weekend transport, approx. 300 means (06-2021).

In the event of delays of means of transport, detour routes e.g. due to construction works, transport restrictions due to cultural or other events or any other event causing anomalies compared to the current timetable, this information from the internal IS of the dispatching centre is published both in the form of a non-numbered list directly on the tab next to the map base and at the same time on the web portal idsjmk.cz according to individual lines, see the following pictures.

Fig. 7: Information about current changes in traffic on the map base of mapy.cz.

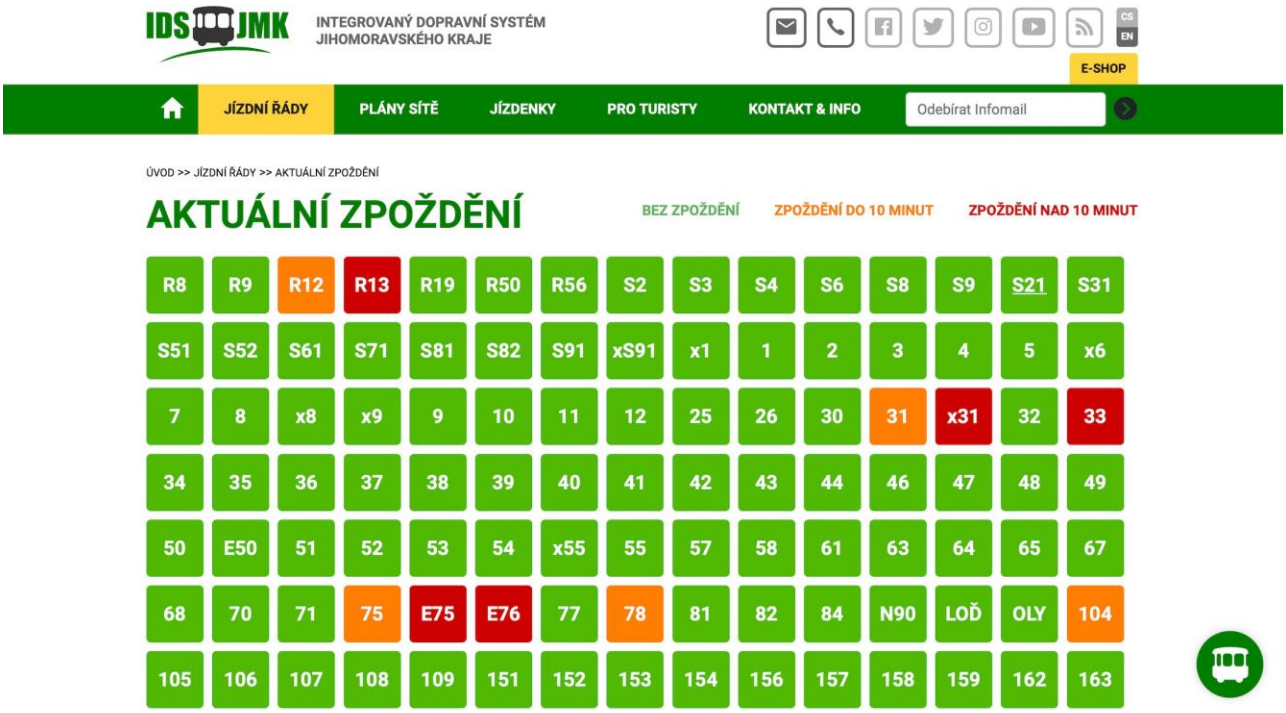


Fig. 8: Information on current delays of individual lines of IDS JMK - overview of all lines.

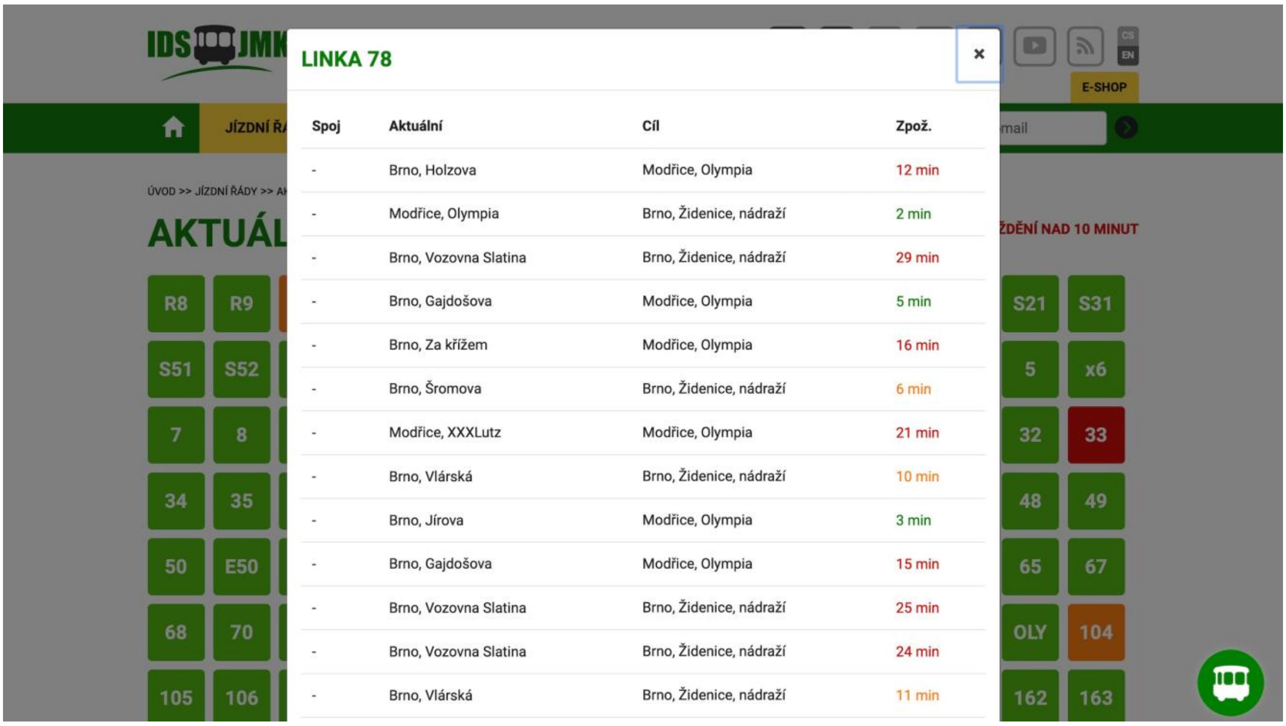


Fig. 9: Information about current delays of individual lines of IDS JMK - delays of selected line.

2.1.2. Transport performance of IDSJMK

In terms of the number of passengers carried, the so-called transport performance is generally monitored in public and private passenger transport. This parameter is given in OSKM (passenger-kilometre) units, i.e. the transport of one person over a distance of one km. In addition to the time aspect, transport performance in the IDS JMK is monitored mainly by individual regions, see the following graph. It can be seen that in addition to the increase in total transport capacity due to the connection of new regions during the aggregation of the IDS JMK in 2004-2010, there is also a slight increase in total transport capacity, especially in the city of Brno and its surroundings.

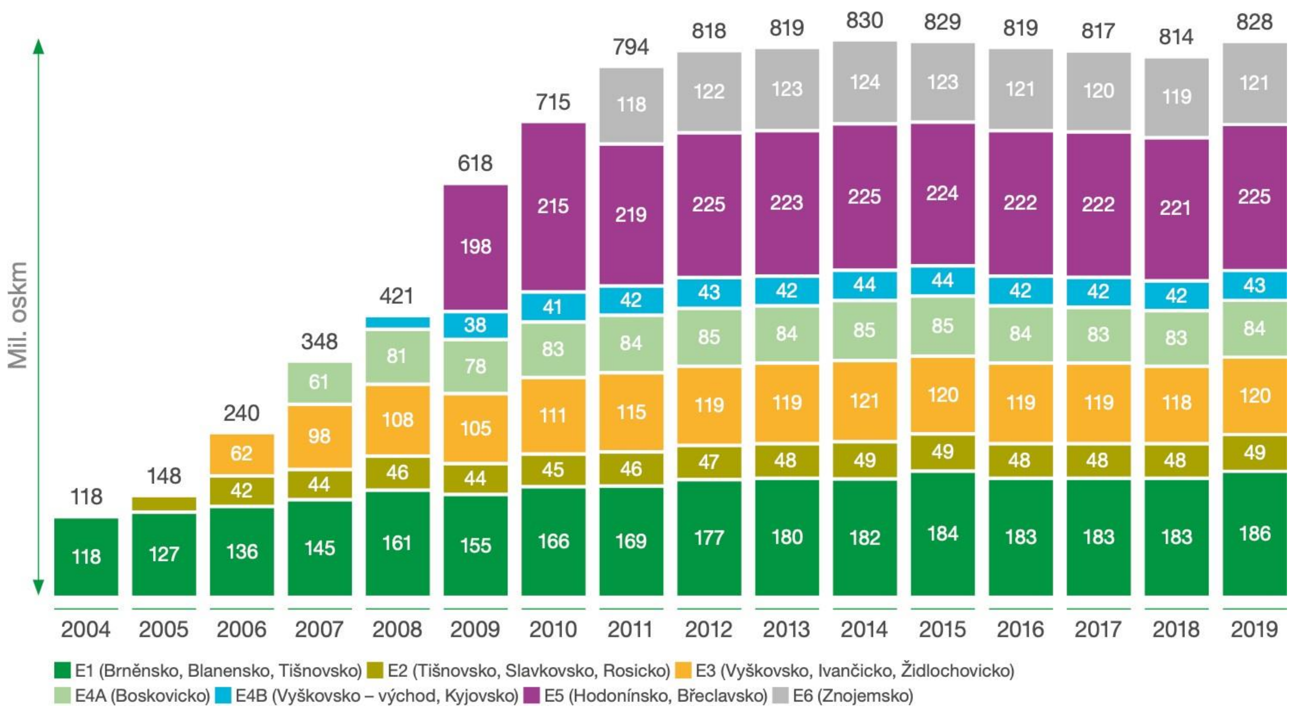


Fig. 10: Transport performance in passenger-kilometres by year and region [23].

2.2. Data for traffic control *in* the JMK IDS

The basic data bases for the JMK transport management and other analyses including anomaly detection are the plan of regional lines of IDS JMK, the network of daily transport lines of the city of Brno and timetable data files:

- Plan of the regional lines of the ISJMK : https://content.idsjmk.cz/mapa/Plan-site- whole.pdf, see also Annex 9.1 for the current status.
- Network of daily transport lines of the city of Brno: https://content.idsjmk.cz/mapa/Plan-site- Brno-den.pdf, see also Annex 9.2.

- Timetable data: the GTFS format (General Transit Feed Specification, [format specification on TransitWiki]) and the proprietary format of Kordis a.s.

In the dispatching process, the current location of the vehicles (both regional and urban) is obtained from the on-board GPS data in the vehicles. From the point of view of data processing for anomaly detection, the type of GPS link placed in the vehicle is not relevant, because from the control system point of view, each GPS behaves as an identical unit due to the uniform interface or wrapper used.

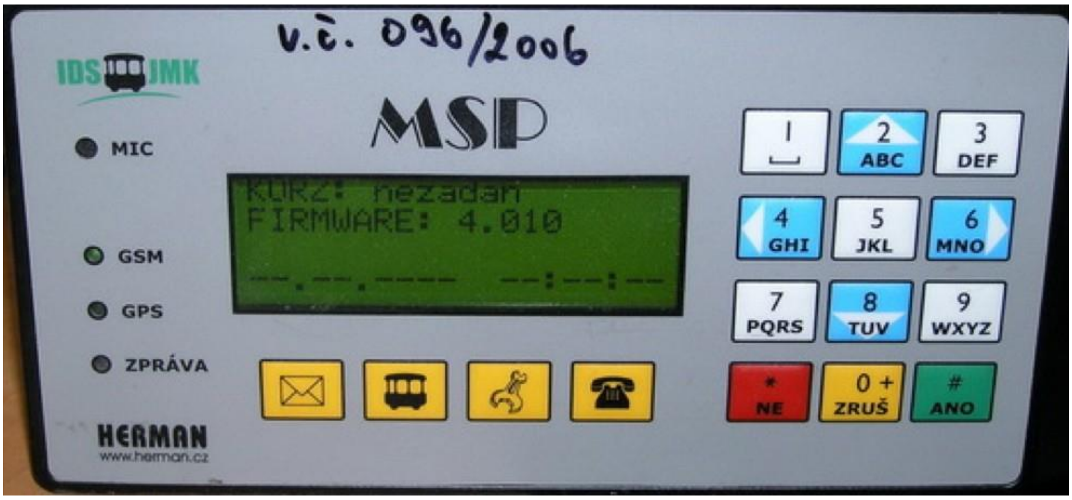


Fig. 11: MSP (Mobile Position Monitoring) terminal of a regional bus [22].

As mentioned in the previous text, the refresh rate for detecting the position of an individual transport vehicle is 6 seconds. In terms of the accuracy of the vehicle positioning, due to the use of links in the civil mode, the accuracy is set between 30 cm and 5 m in the case of the use of the US GPS structure, or 1 m in the case of the use of the European Galileo system. These limitations have to be taken into account as a systematic error of positioning and latency at the input of the data processing system.

3. DEFINITION OF THE ANOMALY DETECTION AND LOCALIZATION TASK

An anomaly, or more technically, a statistically significant deviation, is defined in data processing as any unusual sequence or pattern in the so-called data corpus. A data corpus is a set of data that may have a known or unknown format. The search anomaly may then have a known or unknown structure. The two parameters mentioned above then determine the four basic types of anomaly detection tasks:

1. Structured anomalies in a known data corpus.
2. Structured anomalies in an unknown data corpus.
3. Unstructured anomalies in a known data corpus.
4. Unstructured anomalies in an unknown data corpus.

The role of anomaly detection in the public transport dataset of IDS JMK may correspond to the first and third items of the list depending on the type of anomaly searched for. The first point corresponds to the situation when known (defined) anomaly structures are searched for, here specifically time and position deviations of GPS vehicles in a known data corpus format (e.g. the mentioned GTFS). The third point corresponds to the situation where any anomalies or generally unknown patterns are searched for in that known data corpus without their structure being defined in advance.

The output of the anomaly detection algorithm can take several values depending on the presence of the anomaly and whether the detection is correct:

True positive: correctly detected anomaly

False positive: anomaly detected in error (false alarm) True negative: normal value detected correctly

False negative: a false detection of a normal value (false quiet)

For all detection mechanisms, regardless of whether they use traditional data processing methods or AI or machine learning methods, reducing the sensitivity of the algorithm can reduce the number of false alarms but increase the number of false misses. These two values are always correlated within a given algorithm setting (but not necessarily in exact proportion).

3.1. Anomaly detection from positional *data of* public transport vehicles

As mentioned in Chapter 2, in the current state the location of public transport vehicles of IDS JMK is monitored using GPS links and in case of deviations from the established timetable the calculated delays are publicly displayed numerically in minutes on the web portal idsjmk.cz. In practice, the detected delays are the result of a number of different influences, e.g. delays in the car itself, accidents, congestion, changes in transport organisation, mass sporting or cultural events and other critical or otherwise unpredictable situations.

In addition to displaying the detected anomaly publicly on the website or in the IDS JMK passenger application as an emergency event, a second branch should also be considered for the needs of traffic management, namely the API area of the IDS JMK control room, which will display the detected anomaly and enable the operator to take effective action to correct the situation. In this case, it is advisable to attach additional information to the detected anomaly to facilitate or increase the effectiveness of the operator's intervention in traffic.

This information, essential for effective operator decision-making, can be extracted from the timetable data corpus augmented with the history of public transport operations. It is thus possible to distinguish between situations where the detected anomaly is, for example, a seasonal issue in an annual, weekly or other cycle or, on the contrary, a completely isolated isolated situation to which the dispatching operator has to react personally. It is worth mentioning that, in general, even this intervention of the operator over the anomaly detection system can be replaced by an expert system, but its efficiency is in direct proportion to the volume of accumulated traffic history data and in the case of a completely new type of event, the system will not be able to find a solution, or will use the standard so-called zero solution for unknown situations.

In a first approximation, the detected anomaly, i.e. differences between timetable data (e.g. from GTFS files) and data from GPS links of public transport vehicles, can be classified according to justified generalizability into one of the following categories, namely an anomaly valid for:

a) only the specific car of the corresponding line (isolated car problem),

b) all cars of the corresponding line (isolated line problem),

c) all cars of the selected lines in the affected area (selective area problem),

d) all cars of all lines in the area concerned (general problem of the area).

Each of these variants can then be further expanded, depending on the time window of its validity, into either a one-off or a generally recurring anomaly, either regularly cyclic or randomly acyclic. This results in a set of eight basic types of anomalies for which defined or recommended procedures for dealing with the causes can be provided. For a more detailed specification of the anomaly detection concept in the IDS JMK, see Chapter 4.

3.2. Visual analysis of the traffic situation

Recognition of the cause and type of anomaly can be supported by a superstructure in the form of a CCTV surveillance system with manual or automated congestion detection directly from traffic images. A sufficiently dense camera network in the area of interest is a prerequisite. Benchmarks in the form of datasets containing various annotated traffic situations are then used to analyse the behaviour of the traffic situation recognition algorithms. Examples include the GRAM Road Traffic Monitoring dataset consisting of three types of traffic scenes or the Trafficdb dataset, in which images are divided into Light, Medium and Heavy classes according to the current traffic density.

Figure 12: Example of the GRAM-RTM dataset for the traffic situation recognition test (sunny, fog, intersection).

Recognising the general traffic situation from CCTV systems under different weather conditions is a technologically and knowledge intensive discipline. One of the congestion detection systems from traffic system cameras is [7]. The current traffic density level is detected from the image by a combination of the following attributes:

- the total number of vehicles detected in the image,
- traffic speed (average of detected vehicle speeds),
- traffic flow (dynamic difference in the number of vehicles entering and leaving the image).

Fig. 13: Change in traffic volume in a short time [7].

3.2.1. Example of congestion detection in an image using AI

In the task [7], the authors used an AI system based on deep learning to detect traffic congestion, specifically testing four different models: a cascade of Haar signs and three types of deep learning based detectors (SSD, YOLO and R-CNN). The reported accuracy of vehicle detection on GRAM-RTM dataset images is at 80%, which is a good result for a general traffic scene.

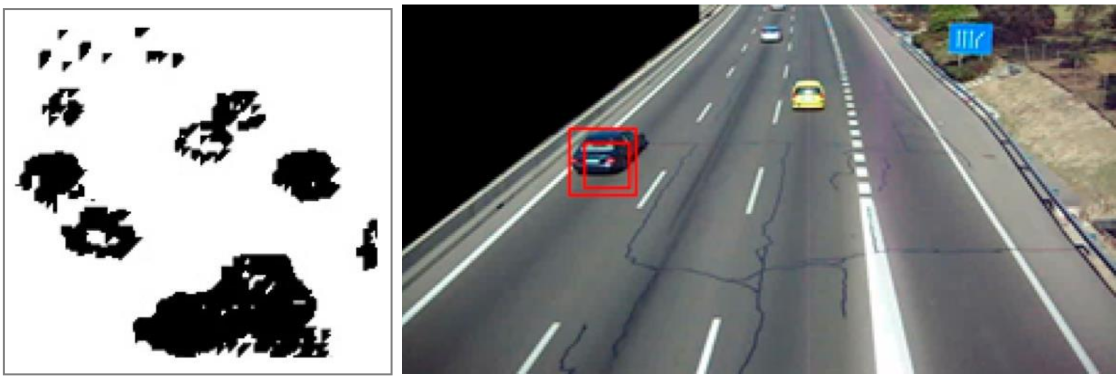


Fig. 14: binary map of detected objects (left) and marking the detection result in the image (right).

For further analysis of this type of congestion detection, the following table comparing the four models on all three parts of the GRAM-RTM dataset (see above) can be used. For each case, the table shows the inference time, i.e. the time required to process one image in terms of classification, and the accuracy of determining the total number of detected vehicles in the scene.

Table 1: Comparison of Haar, SSD, YOLO and R-CNN models for congestion detection.

3.2.2. Example of congestion detection in an image by traditional image processing

An alternative method, until recently prevalent in image processing, is to use one of the traditional image processing methods to detect congestion in a traffic scene. For example, in the task [8], the authors used image texture analysis, which for a traffic scene with sparse traffic is significantly different from a traffic scene containing a high number of vehicles. Vehicles are repeated as objects in the image and thus, from the point of view of texture analysis, it is a repeated image pattern increasing the value of the extracted feature.

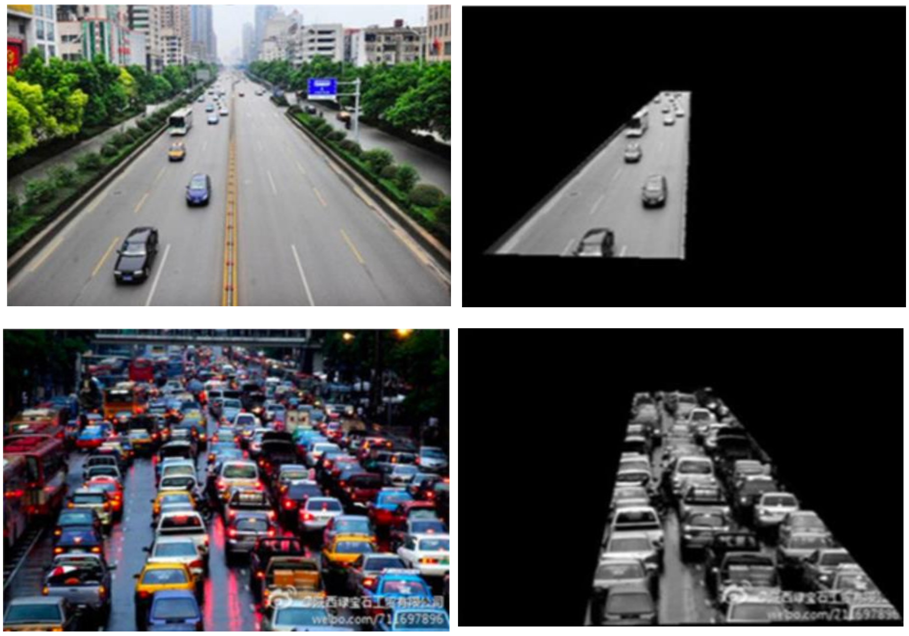


Fig. 15: Statistical analysis of texture in the image with weak (top) and strong (bottom) traffic intensity [8].

The most common method of texture analysis is the use of second order GLCM (Gray Level Cooccurrence Matrix) statistical analysis. The GLCM is always a square matrix expressing the frequency of occurrence of given combinations of pixels (pixels) in an image, thus different statistical results can be obtained for images with different traffic intensities. The energy and entropy features are extracted from the GLCM matrix values, which are used for the subsequent classification of the image in terms of determining the actual traffic intensity level. Computationally, an image period of 6 ms was achieved in the tests, thus the method is suitable for real-time applications.

3.3. Examples of using AI to detect anomalies in public transport data

3.3.1. Singapore - travel time by bus

One of the most efficient traffic management systems is being implemented in Singapore, also one of the world's most congested cities. The National University of Singapore proposed to solve the congestion problem by using an AI-based realistic traffic simulator [14]. The data used for the simulation came from the Open Maps Weather API, Google Maps and the local Land Transport Authority of Singapore. For each registered congestion, geographic coordinates, day of the week, time, current weather, and congestion type were recorded (6 types defined in total). Based on these data, an artificial neural network (ANN) of the MLP-RL (Multilayer Perceptor - Linear Regression) type was learned, the output of which is a prediction (sometimes also an estimation or estimation) of the duration of that particular traffic congestion, more specifically the duration of the trip between two stops in minutes affected by that particular congestion. For each bus route, all congestion estimates along the route are then summed to give an estimate of the duration of the journey along the entire length of the route.

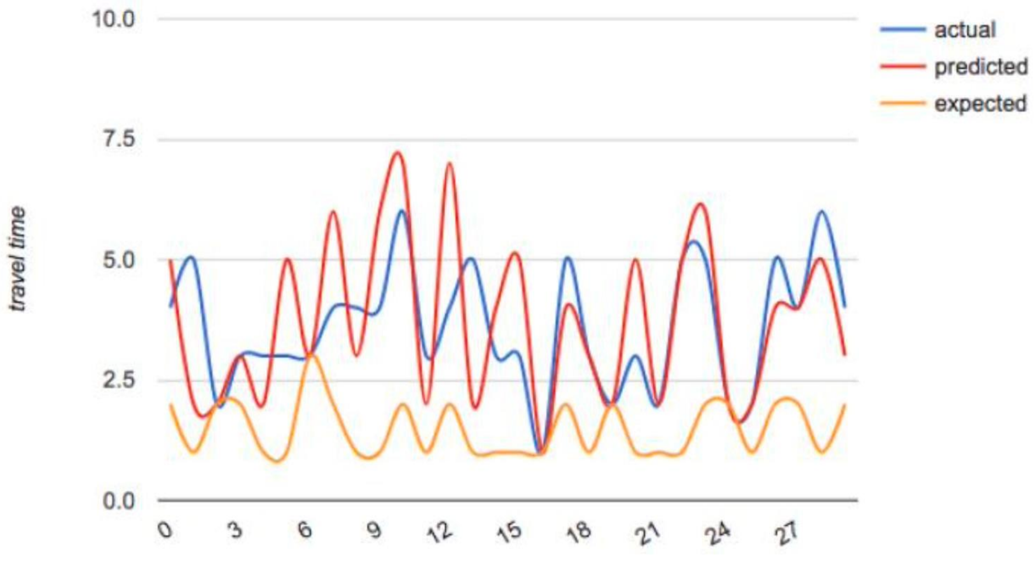


Figure 16: Bus travel time actual (blue), predicted by the simulator (red) and predicted by the transport authority (yellow).

From the above graph for 30 random samples (traffic situations), it can be seen that the ANN model prediction better matches the actual bus journey time than the manual estimation of the transport authority based on experience. Numerically, the increase in the accuracy of the trip duration prediction was quantified at 13%.

3.3.2. Mumbai - travel time by car

At BRAC University in Bangladesh, similar to the previous case, research was conducted to estimate the travel time, but here for a passenger car [15]. The data for training five different AI models (namely: decision tree, random forest, support vectors, linear and logistic regression) were obtained from Uber Movement service which divided the city into nodes. An important effect of this study is the confirmation of the non-negligible effect of weather, and given the research location, the season (monsoon), on the prediction of trip duration.

Table 2: Comparison of the accuracy of the selected AI models on the travel time estimation task.

	Percentage of Data in Test Set			
	40%	50%	60%	80%
Algorithm Name	**Accuracy**			
Decision Tree Regression	42%	54%	68%	73%
Random Forest	56%	62%	74%	83%
Multivalued Linear Regression	48%	69%	75%	84%
Logistic Regression	48%	63%	78%	85%
SVM	35%	60%	68%	70%

The weather data was streamed from the commercial Wunderground Weather service in real time, but was not used as another input for training the models, but as a comparator to determine the correlation between weather type and travel time between city nodes. The weather parameters observed included temperature, humidity, pressure, precipitation, and others, but only the latter was observed to have a statistically significant correlation, namely a 12.6% increase in car travel time between city nodes for every 2.5 cm of precipitation observed.

3.3.3. Beijing - Passenger Forecast

A system for estimating the number of passengers at public transport stops has been developed at Beijing Jiaotong University in Beijing [16]. Here, the AI model is a three-layer ANN, whose inputs are the number of passengers, at a given time of day and day of the week, obtained through an electronic gate at each analyzed stop. The prediction is then made for day *n+1* based on the data of the previous *n* days.

Table 3: Actual (left) and predicted (right) passenger counts for two selected stops of Beijing GUOMAO Station and ANHEQIAO Station in 2018.

Date	GUOMAO Station		ANHEQIAO North Station	
	Actual Value	Predicted Value	Actual Value	Predicted Value
2018/12/17	77910	77372	11553	11530
2018/12/18	78050	78366	11777	11619
2018/12/19	78683	78262	11490	11412
2018/12/20	77921	77458	11478	11551
2018/12/21	80860	81248	12249	12272
2018/12/22	43088	43093	7091	7029
2018/12/23	41197	40167	6263	6334

The training phase used data collected from the electronic gates between 3 September and 13 December 2018. That ANN model exhibited an error in passenger count determination of less than 0.1% after 70 iterations (note: an iteration is a volume unit of learning for an AI model, similar to how a lesson is a unit of learning time in school). In this way, the model was built for 391 different stops and stations of Beijing city.

3.3.4. Stockholm - collective delay detection

The paper [17] deals with the detection of anomalies in the form of unexpected traffic congestion observed on Keolis city buses in Stockholm. Using GPS links, the location and departure and arrival times at bus stops are known for each bus, from which the average speeds of the bus line in different sections of the route and also the number of passengers from the time spent at the bus stop are expertly estimated.

The main added value of the work is the focus on so-called collective delays, i.e. the systematic delay of several units, here buses, in succession rather than the isolated delay of one of the bus. This tests the hypothesis whether congestion caused by specific events such as roadworks, sporting events or concerts etc. can be distinguished by machine learning tools from regularly occurring types of congestion caused by e.g. seasons or other cyclical influences. The AI model used here is the so-called LSTM (Long short-term memory network), which is a very specific (in the sense of less common or unusual) type of recurrent artificial neural network. The model is used to predict a time series, which is then compared with the moving median of the actual delay values of public transport vehicles obtained from their GPS links, and then based on one of the three rules introduced in the paper, a collective anomaly is indicated or not.

3.3.5. Shanghai - taxi driving anomalies by

The subject of a rather unconventional work [18] is an objective assessment of the anomaly of a taxi ride after a customer complaint (long duration of the ride, high price, etc.). Taxi drivers are often penalized by the operator for customer complaints, and usually the reasons for longer journey times, higher fares due to variable traffic conditions or unexpected circumstances are not objectively and ad-hoc assessed. The aim of the solution is to detect real traffic difficulties and thus distinguish between forced anomalous taxi behaviour and deliberately anomalous driving in order to deceive the customer. The proposed methodology was tested on a database of 30 million taxi ride records in Shanghai city over a six-month interval.

This system uses a statistical model with defined limits for anomaly inndication. The algorithm of the proposed model evaluates the following defined problem: for a trip from the origin destination A to the destination B, find the probability of choosing route T from the available set of all known routes based on the distribution of time and routes of taxi trips. Given the statistical constraints, decide whether the current trip T is anomalous (out of bounds) and if so, whether it is anomalous by force or by design. A forced anomalous taxi ride is one that follows a route with a high probability of selection, even if it took longer and was calculated at a higher rate because of the current worsened traffic situation. Conversely, trips along routes with a low probability of collection and comparable or even higher duration and rate calculation are identified as deliberately anomalous.

3.3.6. Nashville - traffic anomalies at sporting events

One of the more recent state-of-the-art methods using the AI model DxNAT in the form of a deep neural network is presented in [19]. Its purpose is to identify non-recurring traffic congestion and also explains its cause.

This method first converts road traffic data into so-called TCI images (Traffic Condition Images), which represent a projection of traffic data onto a geographical map (e.g. vehicle speed on individual road sections). In the next step, a so-called data augmentation (artificial augmentation) and subsequent classification of the current TCI image by a convolutional neural network (CNN). Without the aforementioned data augmentation, only 1440 real TCI images would have been used in one day, a value insufficient for training a deep learning model.

The authors of the paper conducted an experiment involving the detection of traffic anomalies caused by football games and traffic accidents in the city of Nashville. For the test verifying the effect of football matches, traffic data for five days without a match being held in the city and two days when a match was held instead, i.e., the expected proportion of a representative dataset, was used as the training dataset. This resulted in high accuracy in detecting unusual traffic situations correlated in time and location with the events hosted (see also the following Chapter 4 for more information about the project).

3.3.7. *Shanghai* - predicting bus arrival with a collaborative model

In order to improve the accuracy of a real-time public transport information system, specifically the prediction of public transport bus arrival, a collaborative model based on cyber-physical systems architecture is proposed in [20]. The data sources in this model are historical data, i.e., records of previous bus trips, as well as static data such as route segment lengths, intersection traffic light cycles, etc.) and other real-time data from on-board devices. Prediction is performed by fusing all these data sources. A bus line in downtown Shanghai with suitable parameters (route length, number of stops, expected congestion) was selected to verify the method. The results of this work predictably show a synergistic effect, i.e., the above collaborative model achieves higher accuracy than the competing methods presented in the introductory part of the paper [20].

4. CONCEPT OF ANOMALY DETECTION SYSTEM IN IDS JMK

A central feature of public transport throughout its existence anywhere is the general increase in the demands on transport in all basic areas: the amount of passengers and freight transported, the safety of transport, the speed and comfort of transport, the impact on the environment (CO_2 emissions, noise, energy consumption), etc. The increase in demands on passenger and freight transport is directly implied by demographic development, the economic development of society and the globalisation of production and services. In terms of public passenger transport, research on the use of AI in passenger and freight transport is mainly motivated by addressing the defined areas [3]:

- Traffic control.
- Transport safety.
- Public transport.
- Urban mobility.

The main task of this study is to identify possible technological solutions in the field of public transport in the IDS JMK using AI, while it is obvious that always the other areas defined in [3] will be synergistically affected.

4.1. Delays of public transport vehicles compared to the timetable

For the purposes of this study, the general concept of delays of IDS JMK vehicles compared to the timetable should be divided into two parts, namely the detection of delays of a specific vehicle (sensory part) and the processing and use of the obtained information about the delay, i.e. the publication of information about the delay, or action intervention in the transport system to correct the consequences of the situation.

The detection of the delay (and theoretically also the advance) of a specific IDS vehicle against the timetable can always be implemented only under the condition of real-time detection of the position of the specified vehicle. In general, the position of a specific car can be determined in two ways:

- external detection (camera, inductive loops, RFID/NFC or other sensor detector),
- internal sensor (global positioning system, local link).

The advantage of external vehicle position detection is the independence of the vehicle function, e.g. in case of a crash, and the amount of extractable information from a single detector used across the board for all vehicles passing through the detector installation site. The disadvantage of external detection is usually a stationary network of detectors with the need for its own infrastructure (portal, power supply, data transmission).

External detection is suitable for a high ratio of the number of network nodes (here the number of cars) to the number of network nodes (here the stops).

An example of visual detection is a camera system for recognizing the line markings of a public transport vehicle [6]. The visual detection method is based on the assumption of the typical characteristics of the identification LCD/LED panels of public transport vehicles, in particular the unique geometric distribution and the set of colors that can be segmented from the rest of the image, e.g., in the HSV color space. Subsequently, the OCR/OCR algorithm can be used to identify the line number and the destination stop/station. The selection of the image parts geometrically corresponding to the identification panel model can be performed e.g. by a cascade of Haar filters. For the purpose of analysis and implementation of OCR/OCV algorithms using machine learning, a number of datasets exist, e.g. the well-known MNIST (handwritten numbers), SVHN (Street-View House Numbers), and the so-called COCO-Text dataset containing images of a general outdoor scene relevant to the case of this study. A visual detection system is generally more complex and infrastructure intensive than simple sensor systems, but consequently provides a significantly higher amount of usable traffic information.

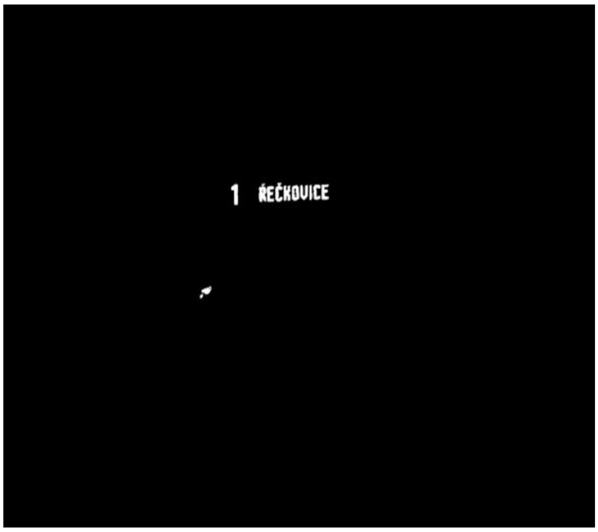


Fig. 17: Example of line number and destination station segmentation using colour space transformation - external vehicle position detection.

The second way to detect the current position of the vehicle is by using an internal sensor installed directly in the vehicle. Due to the mobility of the vehicle, it is obvious that the technology must satisfy the self-localization condition (absolute or relative), regardless of the physical principle used. This could be, for example, satellite navigation modules (Galileo, GPS and other localisation systems) as they are currently used for manual monitoring of public transport vehicles of the IDS JMK.

This method of detection generally appears to be the most effective, both in terms of the available technologies, their parameters, and the complexity of developing the system implementation and subsequent service costs for its operation. As will be seen in the following chapters, this method of locating public transport vehicles is significantly dominant in the applications known to date.

4.2. The concept of anomaly detection in the JMK IDS using GPS

As mentioned in the previous chapters, the JMK IDS vehicles are equipped with GSP links to detect their current position in an interval of about 6 seconds. For the pilot implementation of the system of automatic detection of anomalies in traffic, it is thus proposed to use these links as a basic source of positional information about the vehicles without the need to fundamentally modify or transform the infrastructure of the IDS JMK.

Irrespective of the specific AI architecture subsequently used, the implementation of an anomaly detection system in public transport can be divided into two parts in accordance with the theoretical introduction in the first chapter of this study. The first part corresponds to the learning phase and its task is to set the AI model (e.g., neural network, decision tree, etc.) to a working point by means of a series of high number of traffic examples, both normal smooth traffic and congestion examples of different causes. Schematically, the learning phase can be illustrated by the following graphic, in which the values of Y and Y'' correspond to the correct value determined by the operator (or otherwise by the inputs) or the predicted value, respectively. The value herein may be understood to be either just binary information as to whether or not there is congestion, or a multi-valued or even numeric characteristic indicating the category or severity of the congestion.

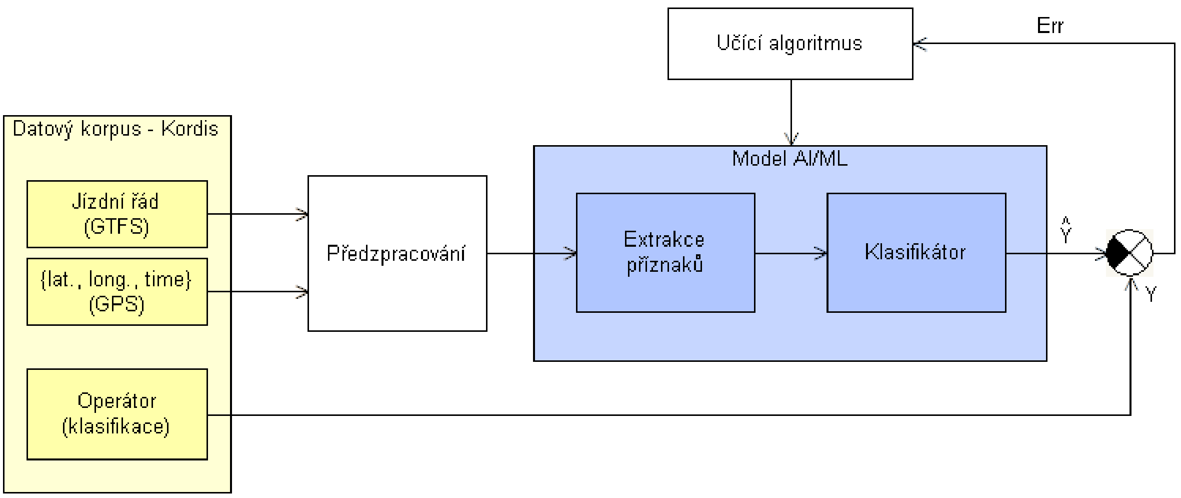


Fig. 18: Conceptual block diagram of anomaly detection using AI incremental learning architecture with teacher-learning phase.

This first, learning phase can take place either in parallel with the operation, from which the input data will only be drawn, but the result of the classification (the so-called prediction) will not enter the public transport control system IDS JMK, or it can take place completely offline on a previously acquired dataset. In the second phase, the so-called classification phase, the output of the learned and verified model is already connected to the IDS JMK, see the following diagram.

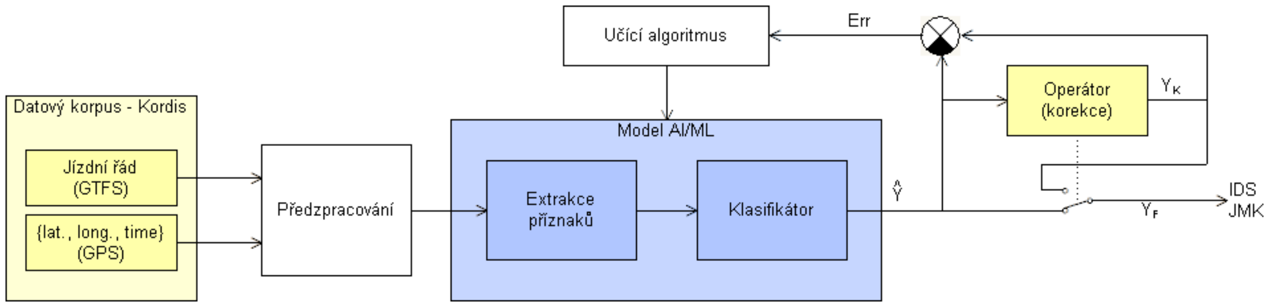


Fig. 19: Conceptual block diagram of anomaly detection using AI architecture of incremental learning with a tutor - classification phase with the possibility of correcting the result (called incremental learning).

The method of connection of the anomaly detector result to the IDS JMK can be selected according to the needs, from its publication only at the operator's workplace for further processing, through automatic display of the result on the public website idsjmk.cz to a fully automated variant with a feedback loop to the IDS JMK control system for real-time congestion compensation, etc. Due to the assumed need to introduce into the prediction system correction information from the operator, which cannot be easily or only with difficulty included in the timetable files, there is also a possibility of immediate correction of the result in the classification phase, with the fact that this correction is also reflected in the previously built model with a certain weight. This feature makes it possible to reduce, for example, the number of repeated congestions of known non-critical cause without the operator having to manually confirm each one.

Several related research works in the field of AI in transport systems, especially from recent years, can be a guide for the implementation of an anomaly detection system in the IDS JMK, regardless of its complexity. Mention can be made of the work [24] from late 2020, in which the authors proposed a system using deep learning architectures (the field of machine learning within AI) that classifies bus trajectories obtained by measuring their GPS coordinates based on the probability of matching (similarity) of the measured trajectory with the trajectory of the planned route of the line. The input GPS data (see the following figure) of the route are merged after preprocessing and form the input of the classification algorithm STOD (Spatial-Temporal Outlier Detector), whose numerical output is the mentioned probability and this, as also implied by the name of the classification algorithm, by using both positional and temporal GPS data of the bus.

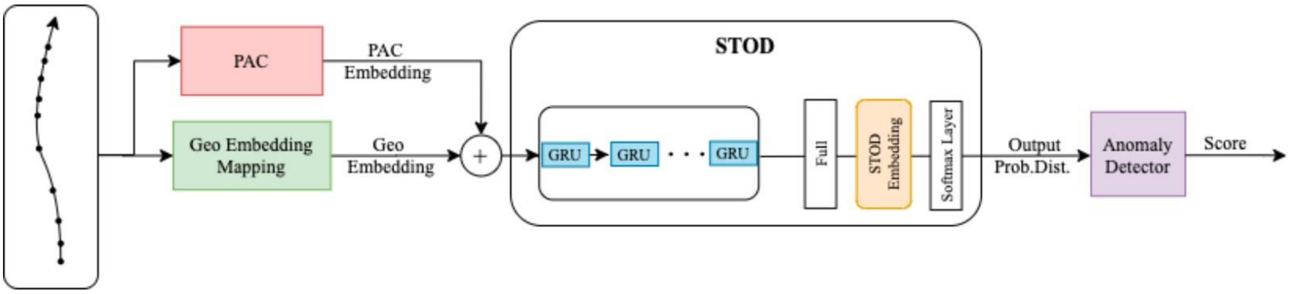


Fig. 20: Process of determining the deviation score of the actual and planned bus route according to [24].

Specifically, the classifier proposed in this paper, using, among others, neural network architecture, determines the so-called anomaly score, whose higher value at the output of the classifier indicates a deviation of the bus trajectory from the expected route, taking into account noise in the data or measurement uncertainty. The higher the value of the anomaly score, the more the measured trajectory deviates from the planned trajectory.

The model is being tested in Recife, Brazil (82 bus routes) and Dublin, Ireland (66 bus routes) for real-time detection of anomalies caused by congestion and detours, which bus drivers choose independently or on the basis of dispatchers' instructions based on current traffic information. The system uses the aforementioned GTFS timetable format together with additional supplementary data from OpenStreetMap and open city data as part of the ground-truth to build the input dataset.

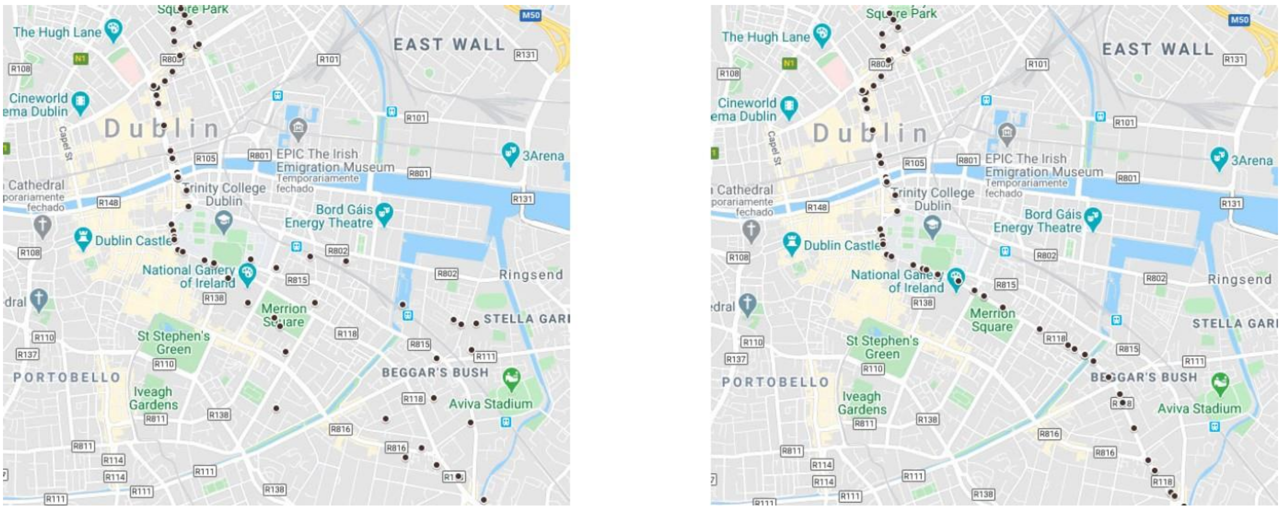


Figure 21: Example of a spatial anomaly (Dublin, Ireland): measured bus trajectory (left), planned route (right) [24].

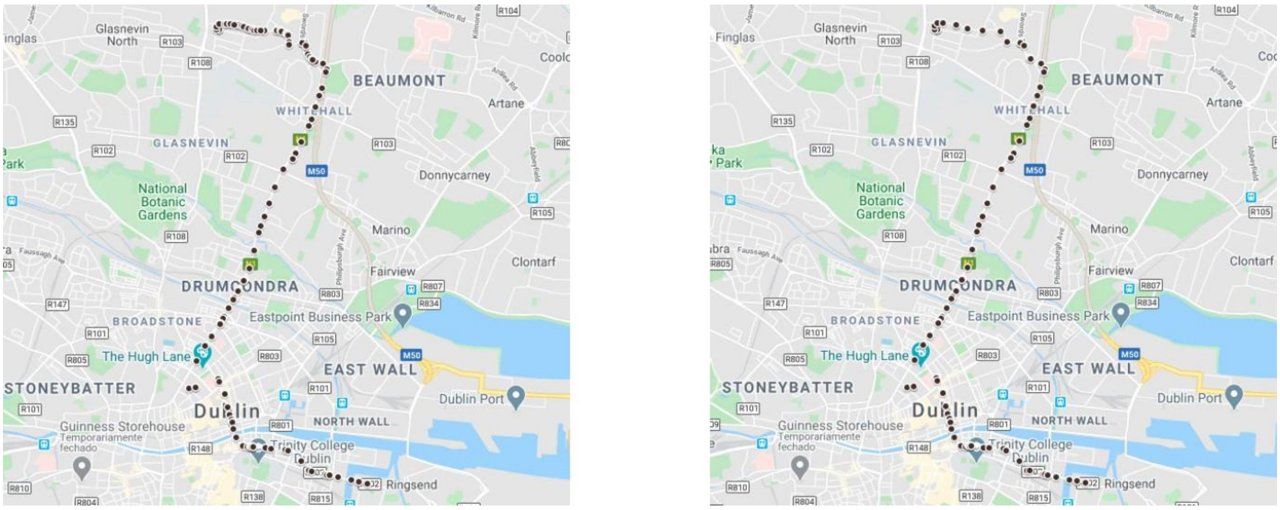


Fig. 22: Example of a time anomaly (Dublin, Ireland): measured bus trajectory (left), planned route (right) [24] - the discrepancy between the measured and planned route timings can be traced from the density of the displayed points, which represent GPS data acquired at equidistant time instants at 30 s intervals.

In [25], in addition to the theory of data management of public transport, contextual prediction mechanisms for short and long term delays in the transit network and optimization algorithms for schedules according to seasonal effects, the approach of using deep neural networks, as a representative of one category of AI methods, for contextual detection of anomalies in the transit network is also developed. A detailed description of the implementation of the method is far beyond the capacity and knowledge requirements of this paper, so only the basic parameters are summarized here:

- common types of causes of congestion are considered: accidents, sporting, cultural and other events, adverse weather, closures, etc,
- The dataset contains over 900 million records and was collected over a period of more than a year in the city of Nashville (a comparable agglomeration to the JMK),
- A multi-layer DNN (deep neural network) is used as a classifier model to search for congestion patterns,
- 98.73% accuracy was achieved in identifying congestion caused by football games with a stadium capacity of 70,000 people.

Note: the accuracy, called Accuracy, is defined in AI as the ratio of correctly classified positive and negative cases to all cases. So the number given means that 1.27% of all cases were misclassified as either a false alarm or a false calm.

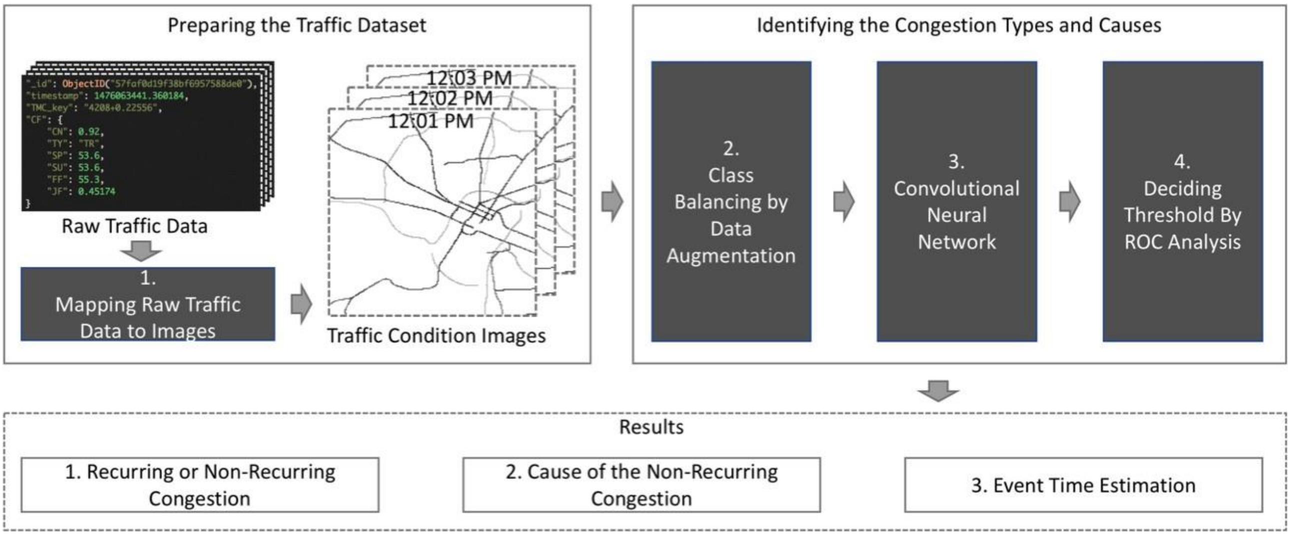


Fig. 23: Block diagram of anomaly detection using DNN according to [25].

The implementation of the method was verified and the stated accuracy obtained in the so-called motivational example of eight football games at Nissan Stadium at the turn of 2016 and 2017. The basic parameters and the impact of hosting the games on the traffic around the stadium can be illustrated by the following graphic.

Date	Start Time (CST)	Stadium	Attendance	Duration (HH:MM)
1/1/17	12:00 PM	Nissan Stadium	65205	3:11
12/11/16	12:00 PM	Nissan Stadium	68780	3:02
11/13/16	12:00 PM	Nissan Stadium	69116	3:36
10/27/16	7:26 PM	Nissan Stadium	61619	3:08
10/23/16	12:02 PM	Nissan Stadium	65470	3:21
10/16/16	12:02 PM	Nissan Stadium	60897	3:12
9/25/16	12:02 PM	Nissan Stadium	62370	3:03
9/11/16	12:05 PM	Nissan Stadium	63816	2:57

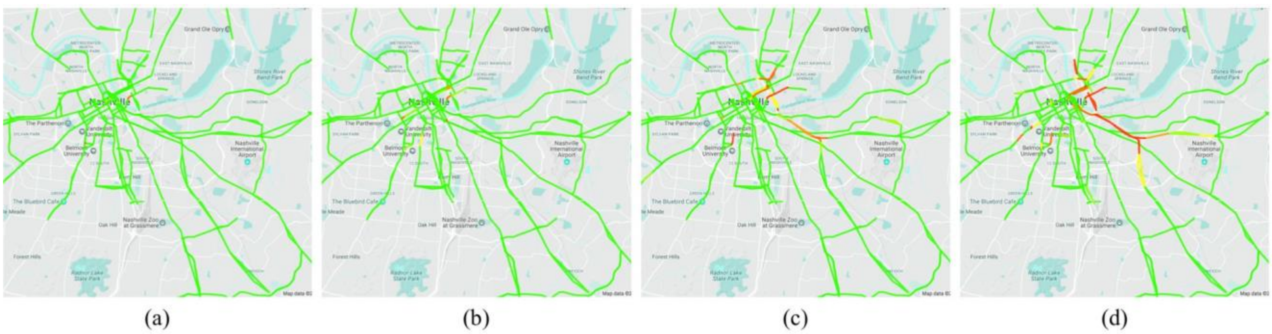


Fig. 24: Tabular breakdown of football games with visitor numbers and duration (top right two columns) and corresponding traffic density around the stadium of four hours (a), three hours (b), two hours (c) and one hour

(a) before the start of the game [25].

There are many similar works within the research of data analytics of public transport control systems using modern tools using AI, Data mining or Big data knowledge, but they are usually not yet implemented in the transport system (more precisely, these implementations are not yet widely published).

4.2.1. Timetables in GTFS

Regardless of the specific GTFS format, this and any other ground-truth for detecting anomalies in public transport in the IDS JMK or elsewhere. It is a basic data set against which the measured actual traffic data is compared and based on the detected deviations it is decided what type or intensity of deviation caused by congestion or other traffic event is involved.

GTFS (The General Transit Feed Specification) defines a general format for public transport timetables and associated geographical information. The GTFS format allows public transport operators to publish transit data in a unified format that enables the development of other transport applications by third parties, such as online route planners, in addition to the implementation of their own management systems. The GTFS specification has a fixed structure and way of working with the information, which is stored in uniquely defined files divided according to the type of information represented as follows:

- `agency.txt`
- `stops.txt`
- `routes.txt`
- `trips.txt`
- `stop_times.txt`
- `calendar.txt`
- `calendar_dates.txt`
- `fare_attributes.txt`
- `fare_rules.txt`
- `shapes.txt`
- `frequencies.txt`
- `transfers.txt`
- `pathways.txt`
- `levels.txt`
- `feed_info.txt`
- `translations.txt`
- `attributions.txt`

Some of the data files listed above are mandatory, i.e. required for the correct functioning of transit applications (e.g. stops, routes, trips, etc., containing basic definitions lines, times, etc.), while other data files are optional (e.g. fare_attributes, transfers, levels, etc.) and are used for superstructure services. Here, e.g., for fare calculation, transfer planning at transport nodes, and information about floor changes at stations. See also Annex 9.5 for a tabulated list of data sets with a more detailed description.

The data in the individual files are interlinked, in particular the files stops.txt, routes.txt and trips.txt have direct bilateral links between them. A specific example of the content of some selected GTFS files illustrates well the philosophy of storing traffic data, which was originally inspired by the general comma-separated value format (the well-known CSV format).

agency.txt

```
agency_id,agency_name,agency_url,agency_timezone,agency_lang
agency001,Transit Agency,http://www.transitcommuterbus.com/,PST,en
[...]
```

stops.txt

```
stop_id,level_id,stop_name,stop_lat,stop_lon,location_type,parent_station
F12,,5 Av/53 St,40.760167,-73.975224,1,
E1,L0,5 Av/53 St SW,40.760474,-73.976099,2,F12
E2,L0,5 Av/53 St NE,40.76035,-73.97546,2,F12
E3,L0,5 Av/53 St SE,40.760212,-73.975512,2,F12
E4,L0,Madison/53 St NE,40.759612,-73.973731,2,F12
E5,L0,Madison/53 St SE,40.759491,-73.973820,2,F12
N1,L1,,,40.760457,-73.975912,3,F12
N2,L1,,40.760531,-73.976111,3,F12
N3,L1,,40.759746,-73.974203,3,F12
N4,L1,,40.759679,-73.974064,3,F12
F12S,,5 Av/53 St,40.760167,-73.975224,0,F12
B1,L2,,40.759746,-73.974203,4,F12S
B3,L2,,40.759828,-73.974442,4,F12S
F12N,,5 Av/53 St,40.760167,-73.975224,0,F12
B2,L3,,40.760457,-73.975912,4,F12N
B4,L3,,40.760375,-73.975729,4,F12N
```

1, ⊙: station; 0, ⊙: platform; 2, ⊙: entrance/exit; 3, ⊙: generic node; 4, ⊙: boarding area

routes.txt

```
route_id,route_short_name,route_long_name,route_desc,route_type
A,17,Mission,"The ""A"" route travels from lower Mission to Downtown.",3
```

trips.txt

```
route_id,service_id,trip_id,trip_headsign,block_id
A,WE,AWE1,Downtown,1
A,WE,AWE2,Downtown,2
```

stop_times.txt

```
trip_id,arrival_time,departure_time,stop_id,stop_sequence,pickup_type,drop_off_type
AWE1,0:06:10,0:06:10,S1,1,0,0
AWE1,,,S2,2,1,3
AWE1,0:06:20,0:06:30,S3,3,0,0
AWE1,,,S5,4,0,0
AWE1,0:06:45,0:06:45,S6,5,0,0
AWD1,0:06:10,0:06:10,S1,1,0,0
AWD1,,,S2,2,0,0
AWD1,0:06:20,0:06:20,S3,3,0,0
AWD1,,,S4,4,0,0
AWD1,,,S5,5,0,0
AWD1,0:06:45,0:06:45,S6,6,0,0
```

levels.txt

```
level_id,level_index,level_name,elevation
L0,0,Street,0
L1,-1,Mezzanine,-6
L2,-2,Southbound,-18
L3,-3,Northbound,-24
```

From the above examples of GTFS files, it can be easily observed that e.g. only from the files stops.txt and stop_times.txt it is possible to detect a possible deviation of the current state of the corresponding vehicle from the timetable by simple differentiation against the data from GPS links (latitude, longitude, time). The GTFS data files, together with other necessary data, can serve as the dataset mentioned in the introduction for anomaly detection using AI architecture, e.g. decision tree, Bayesian classifier or neural network.

4.2.2. Primary anomaly detection

From a practical point of view, it is always necessary to have a pair of values to build an anomaly detection algorithm - the actual value (the ground-truth - here the timetable, specifically e.g. the time data) and the predicted or measured value (here e.g. the GPS data of the car). For the task detection of anomalies in public transport of IDS JMK let us consider in the framework of the system implementation a standard task of the type of learning with a teacher, i.e. to the mentioned pair of values also available information whether it is a case of anomaly or not. This information can be either binary, i.e. only yes-no, or qualitatively and quantitatively scalable according to the severity of the anomaly. Primary anomaly detection then consists of a time-limited learning phase, which was mentioned in the introductory chapter of the study, and a subsequent classification phase, where the detection system no longer needs operator intervention. The deviations of the currently measured traffic values from the valid timetable are evaluated in the familiar way learned in the previous phase. The length of the learning phase always depends on the number of learning examples in it.

4.2.3. Adaptive adaptation of anomaly detection to traffic *changes*

Due to the complexity of the learning phase in terms of the number of representative cases of all anomaly categories (congestion, concert, accident, vehicle failure, traffic volume, etc.) and also e.g. the need to capture seasonal cyclic anomalies with a yearly period (e.g. systematic delay of trainsets due to winter weather conditions in a certain area), it is appropriate to allow further adaptive adaptation of the already learned system to further changes after a limited learning phase for the primary anomaly detection. This "lifelong" biological process, common to humans, is called incremental learning in the case of machine learning, which has its own not entirely intuitive implications and places specific requirements on the AI architecture. In principle, however, incremental learning can achieve corrections and modifications to a system already learned by an operator and thus, for example, correct for seasonal effects over a longer period of time than just one year. **As a result, it is possible to achieve a system capable of extracting intrinsic (not easily observable from the outside) properties of traffic flows and thus classifying detected anomalies both automatically and often in a much more efficient *way compared to* human operators**. This result is due, besides the extracted intrinsic knowledge of the system itself, also to the amount of data processed simultaneously in this way.

4.3. Examples of anomaly detection in transport - overview

In this chapter, examples of selected systems for anomaly detection in urban agglomerations of different types given by the geographical location and size of the area managed by public transport are presented. It is important to note here that at present (2021), the collection of traffic information is mainly automated in traffic management systems, either by external monitoring systems (usually cameras) or on-board sensors, while the actual evaluation of the information obtained is usually performed manually or semi-autonomously, i.e. manually on the basis of autonomously statistically pre-processed data (aggregation, statistical indicators for dispatching operators, etc.). The analysis of

AI approaches in the field of public transport presented in the previous chapters thus implies to some extent, in the case of system implementation, also research activities, apart from development.

4.3.1. Congestion detection using taxi vehicles with GPS

For fast and efficient congestion detection in the Beijing metropolitan area, the authors of the system described in [9] used the current network of taxi vehicles, which are natively equipped with GPS locators. The traffic map of the city displayed on the geographic background is divided into a uniform rectangular grid of square segments. Each segment represents a further indivisible entity for congestion detection in which data from all GPS locators of taxi vehicles within that segment are aggregated.

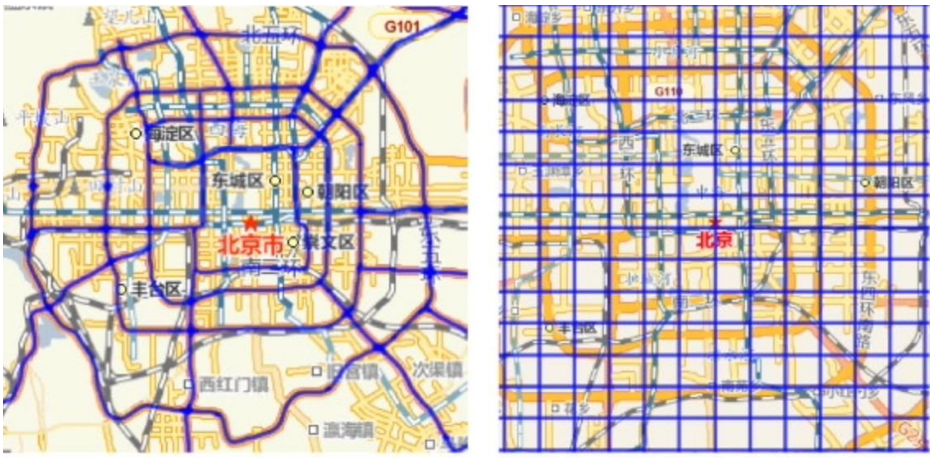


Fig. 25: Traffic map of Beijing city (left) and the division into sectors for congestion indication (right) [9].

A model has been developed for standard traffic behaviour, from which deviations are then statistically evaluated. The model is built on traffic flows characteristic of the segments at a given time, i.e. reflecting the movement of people to and from work at a given time of day, increased or decreased traffic flows to the airport at times of aircraft arrivals and departures, the same for train services, etc. Taxi vehicles, in terms of their number and the size of the area served, are a reliable indicator correlating with actual traffic volumes.

Without prejudice to the accuracy and completeness of the study, it can only be briefly mentioned here that the automatic identification of cells with a statistically significant deviation from the standard (Expectancy = mean value in statistics) traffic is performed in real time with the LRT (Likelihood Ratio Test) over an aggregated map of segments with a known number of currently present GPS links.

4.3.2. Classification of anomalies in the City of Boston traffic dataset

As part of university research at Boston University, a method for classifying anomalies in traffic data was proposed by a pair of authors in [10]. The goal of real-time detection of emerging traffic congestion is early operator (here manual) intervention to mitigate or completely compensating for the extent of congestion. The basic architecture of an anomaly detection system can be illustrated by the following diagram.

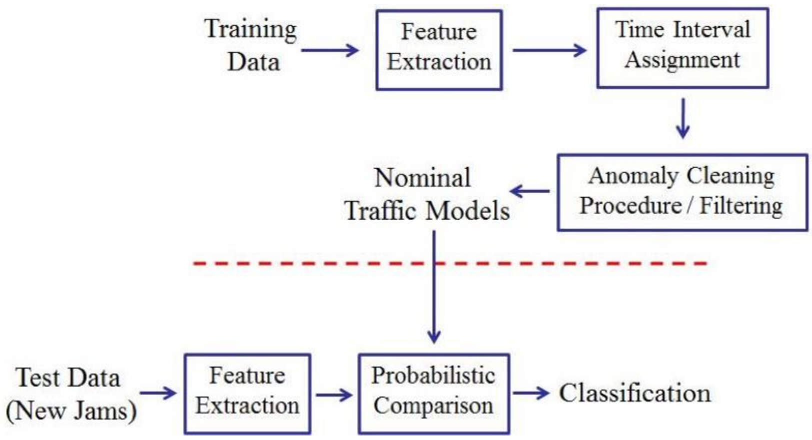


Fig. 26: Architecture of the traffic anomaly detection system of the city of Boston [10].

The input dataset for the research was provided by the City of Boston in the form of a list of all recorded congestion from 2014 to 2016. The detected congestion is divided into its categories in the dataset:

- regular (expected) congestion appropriate to the location and time of day,
- anomalous congestion caused by traffic accidents, reconstructions and other influences.

At the same time, each record corresponding to a single congestion contains ordered information according to the specification:

- time stamp,
- Record ID / congestion,
- geographical coordinates,
- other traffic congestion parameters (extent, duration, etc.).

Based on these input dataset data, the following four parameters (reduction of the feature space, i.e., the amount of data extraction relevant) are calculated for each record:

- average latitude (*lat*),

- mean longitude (*long*),

- congestion orientation, i.e. the angle relative to the positive part of the horizontal axis of the map (),

- congestion length (*l*).

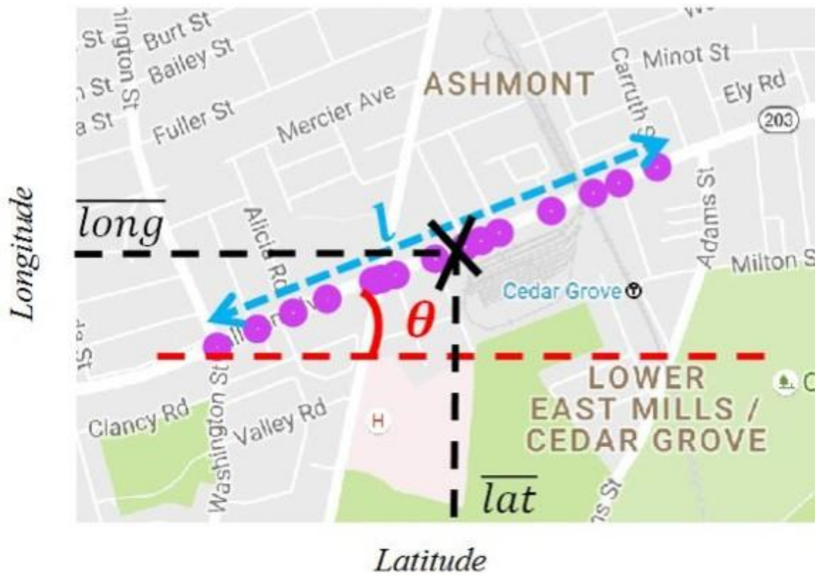


Figure 27: Illustration of extracted features (coordinates, orientation and length) for classification of the degree of anomaly of the detected congestion.

The classification of congestion based on the four extracted features takes into account both the standard model of the ideal traffic situation and the regular (expected) deviations in normal traffic. The expected deviations are mainly caused by the periodicity of the seasons (different traffic behaviour in winter and summer, holidays, etc.), the weekly working cycle (weekdays, weekend, holidays) and the daily intervals (e.g. morning and afternoon peak hours). For each such period, a model of expected traffic behaviour is then created, against which the new traffic data is then compared using the clustering methods DP-means and DB-scan and finally classified according to the distance in 4D space from known clusters representing known congestion or a standard traffic behaviour model. An example of the evaluation of several new situations is given in the following summary table.

Table 4: Table of symptoms (first four columns) and corresponding congestion classification (last column).

Test Results for Real 4D Data - DP Means + DB SCAN				
Longitude	Latitude	θ	l	$p_{association}$
-71.1022	42.3494	-35	.6	0.8570
-71.1022	42.3494	-35	1.5	0.6725
-71.0873	42.3464	-60	.3	0.8707
-71.1715	42.3331	-60	.3	0.6699
-71.1072	42.3389	45	.48	0.8616
-71.1122	42.3239	-79	.28	0.6979
-71.2056	42.3947	-40	1.1	0.5806
-71.2524	42.2127	20	.34	0.4625

The value of the classification parameter in the last column ($p_{association}$) indicates the probability with which the congestion corresponds to a known model of the usual traffic situation. Thus, a high value indicates congestion that is known and therefore expected at a given location and time of day, and conversely, a low value of $p_{association}$ indicates an anomaly relative to both normal standard traffic and an anomaly relative to expected deviations at a given time of day at a given location.

4.3.3. Anomaly detection from GPS data of public transport buses

Systems based on tracking the GPS position of taxi vehicle locators are popular and therefore common in the development of anomaly detection systems. Nevertheless, anomaly detection solutions based on similar processing of the same GPS link information, but from public transport vehicles, are emerging as a more suitable solution for traffic management by urban enterprises. This compensates for some influences such as the route preference of taxi drivers and the systematic offset given by the target group of taxi users. An example of a proposed method using deep learning architectures can be found in [11] (Australia/China - city of Kuej- yang, approx. 5 million inhabitants in a mass transit metropolitan area).

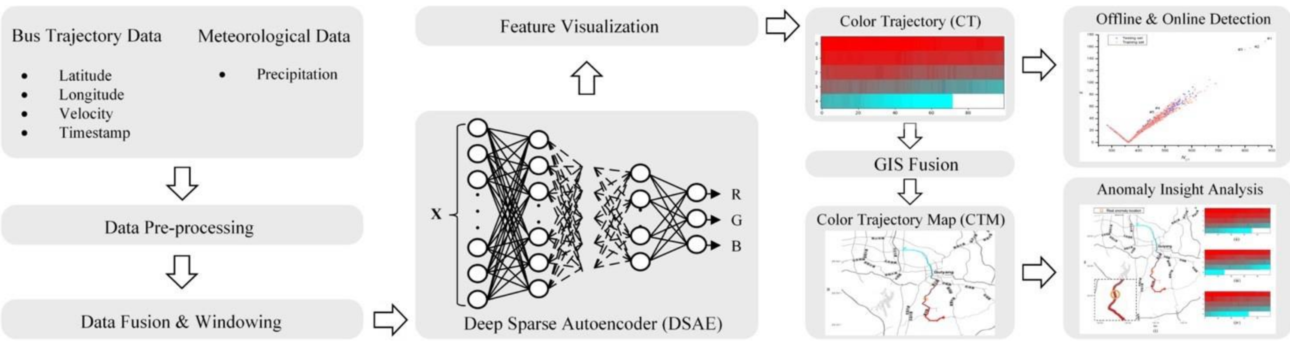
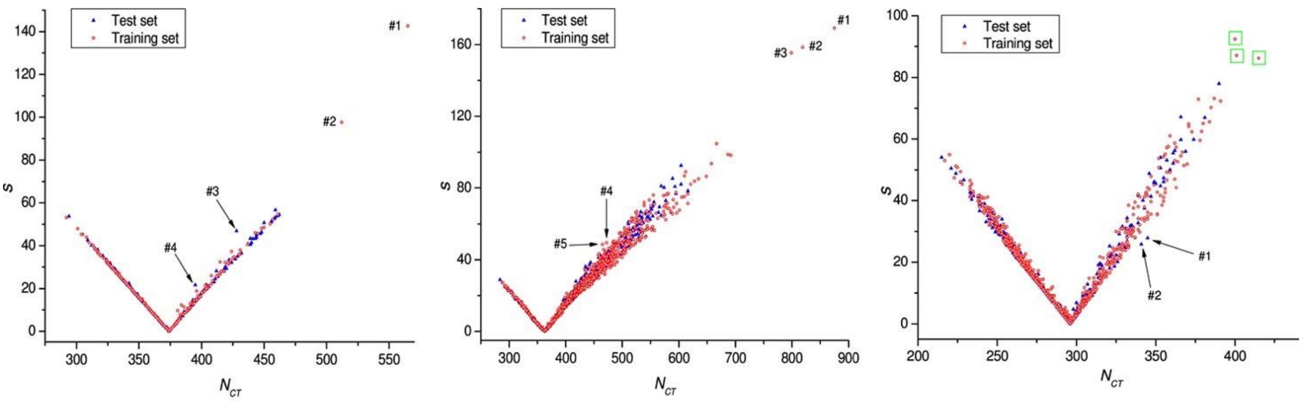


Figure 28: The process of traffic anomaly detection using deep learning [11].

The scope of the study does not allow for a detailed description of the principle of deep learning architectures, therefore it will be limited here to stating that it is a computational model whose input is raw ranked data of a traffic situation and its output is extracted features relevant for the recognition of the type of traffic situation. It should be emphasised that the classification capability of the model is dependent on the accuracy of its setup in the learning phase (see introductory chapter). The previous figure shows schematically the anomaly detection process according to [11]. The computational block of deep learning, which is a key component of the system, is represented in the diagram by a network of links connecting the nodes of the network (note: technically correct here is an artificial neural network with neurons arranged in layers from input to output).

In the case of deep learning architectures, knowledge about congestion is encoded in the internal weights of the network. The advantage of these approaches is their applicability even for non-explicitly defined tasks. A related disadvantage is the lack of the ability to extract relevant congestion knowledge from input data, in other words, it is not easy to trace on the basis of which procedure a given input data was classified into a given class (e.g., congestion of a certain type and intensity).



(a) Bus Route 66 **(b)** Bus Route 50 **(c)** Bus Route 18

Fig. 29: Plots of two selected attributes ($_{NCT}$ and s) for the three studied bus routes 66, 50 and 18 - the individual colored points correspond to the observed public transport vehicles (red points to the training data, blue points to the test data) - the anomalies can be traced as points distant from the clusters.

The format of the resulting classification of detected anomalies into two defined classes A and B can be shown in the following table. Each detected public transport bus anomaly is assigned an identifier, a time stamp and a classification of the event and its placement in one of the two classes.

Table 5: Output of the anomaly classifier for bus lines in the Chinese city of Kuejyang.

Route	No	Service date	Running time	Event	Category
66	**#1**	18/09/2016	07:30–09:01	Event 2	Class A
	#2	18/09/2016	07:00–08:23	Event 2	Class A
	#3	26/11/2016	12:43–13:55	Event 3	Class B
	#4	26/11/2016	12:07–13:09	Event 3	Class B
50	**#1**	18/09/2016	06:58–09:25	Event 2	Class A
	#2	18/09/2016	07:22–09:40	Event 2	Class A
	#3	18/09/2016	07:34–09:48	Event 2	Class A
	#4	14/08/2016	19:41–21:01	Event 1	Class B
	#5	14/08/2016	17:32–18:51	Event 1	Class B
18	**#1**	14/12/2016	09:31–10:28	Event 4	Class B
	#2	14/12/2016	09:50–10:48	Event 4	Class B

In the previous table, the anomaly classes are defined according to the degree of deviation in the spatial and temporal domain (corresponding to the two selected NCT and s symptoms in the graphs above), with class B being a less severe anomaly than class A, which corresponds to the anomaly of the observed car in terms of both spatial and temporal symptoms. Simply put, the observed public transport car classified in class A is off-route (detour, reconstruction, accident) and moreover at the wrong time (delays e.g. passenger boarding, etc.).

4.3.4. Estimation of tram delays in the Warsaw transport system

In a 2017 paper [12], authors from Warsaw University of Technology focused on real-time estimation of tram delays from freely available data from the Warsaw Transport Authority. The data is open as a streaming stream through a web interface with the address https://api.um.warszawa.pl. It contains information on current tram locations and timetables and is also used by freely available route planners such as Google Directions for adaptive route planning based on the current traffic situation (so-called situation-aware routing).

Based on the available data, the authors proposed a method to calculate the current delay estimate for each single tram car. This data is then also used to calculate a global indicator of the level of traffic intensity at a given location - essentially a generated heat-map indicating the current traffic density. The architecture of the tram delay estimation system itself is illustrated by the following diagram.

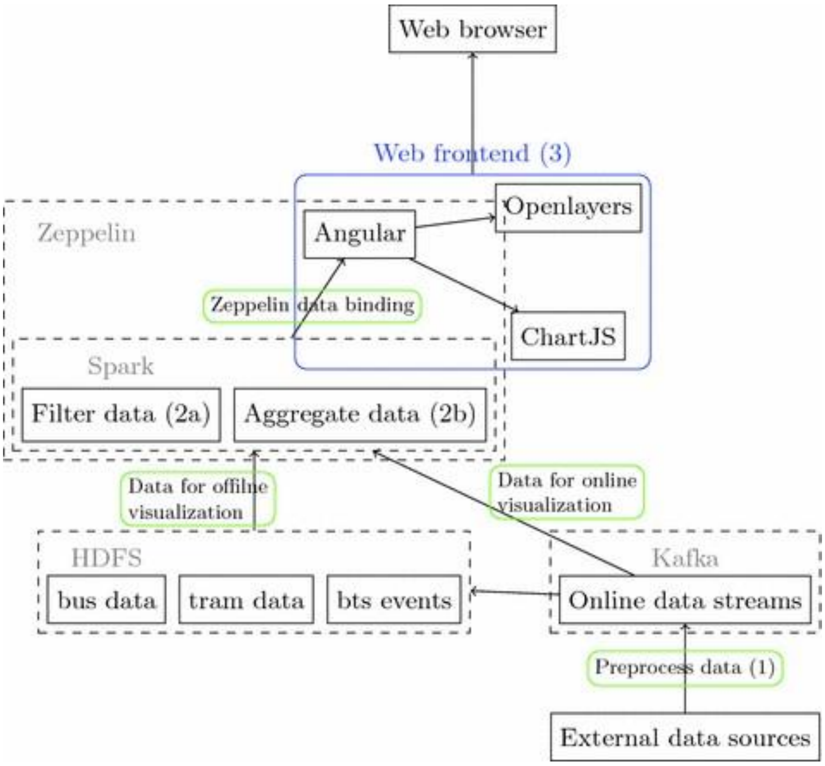


Fig. 30: Architecture of the system for estimating tram line delays in the city of Warsaw [12].

The system analyses the available traffic data for a total of 750 trams, which are updated in a 10-second cycle and represent a data flow of approximately 72 MB/hour. The estimated aggregated delay at a given map location is displayed by the system in real time and the following graphic shows, among other things, that increasing traffic volume can be effectively estimated in units of hours.

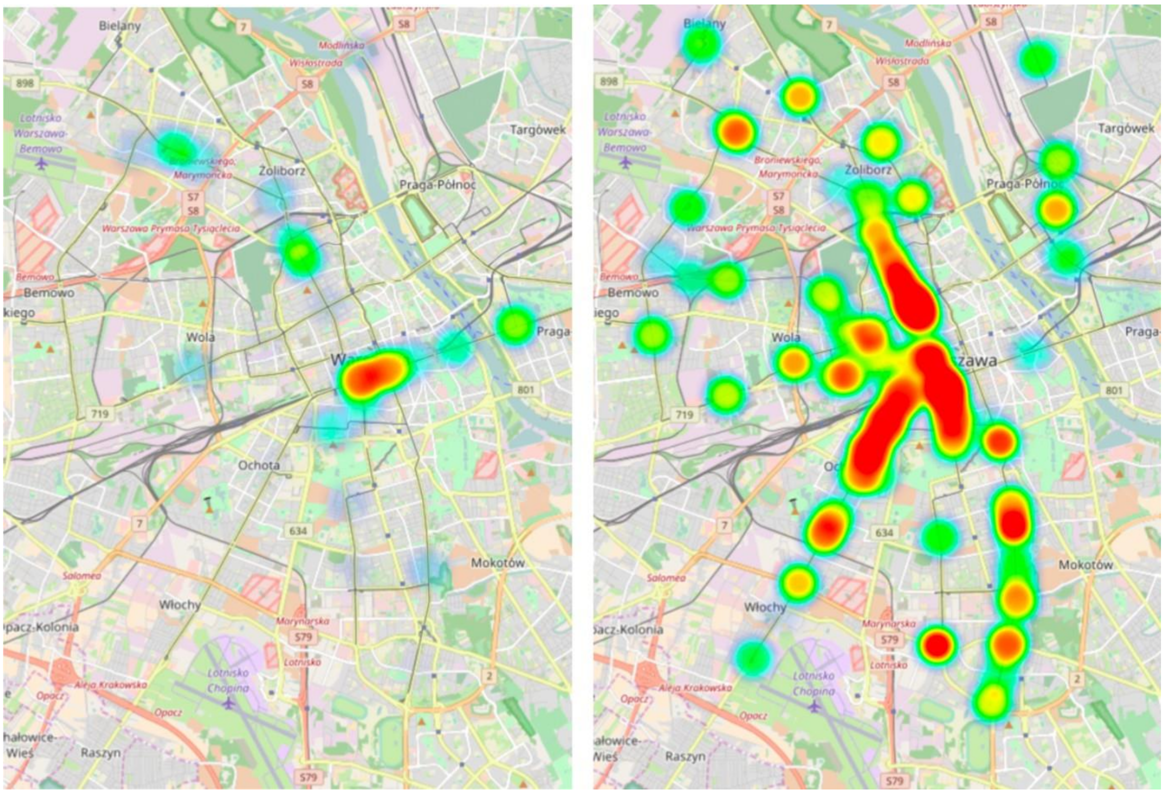


Fig. 31: Detection moment of the excess traffic volume level in the city centre of Warsaw at 16:57 (left) and the absolute peak level (peak-point) in the same area at 17:57 (right).

The estimation of the actual delay of individual tram cars is itself made by a non-trivial calculation. This involves both a comparison of the timetable of a given stop with the actual time of the tram car according to the on-board GPS data (this delay is important especially for passengers at subsequent stops) and an estimation of the actual delay between neighbouring stops (this delay is important especially for the estimation of traffic density for route planners, as it takes into account e.g. traffic preference by switching traffic lights etc.). The result of the detected delays is aggregated for all trams and graphically displayed in a geographical map of the city, where each tram has its own line number.

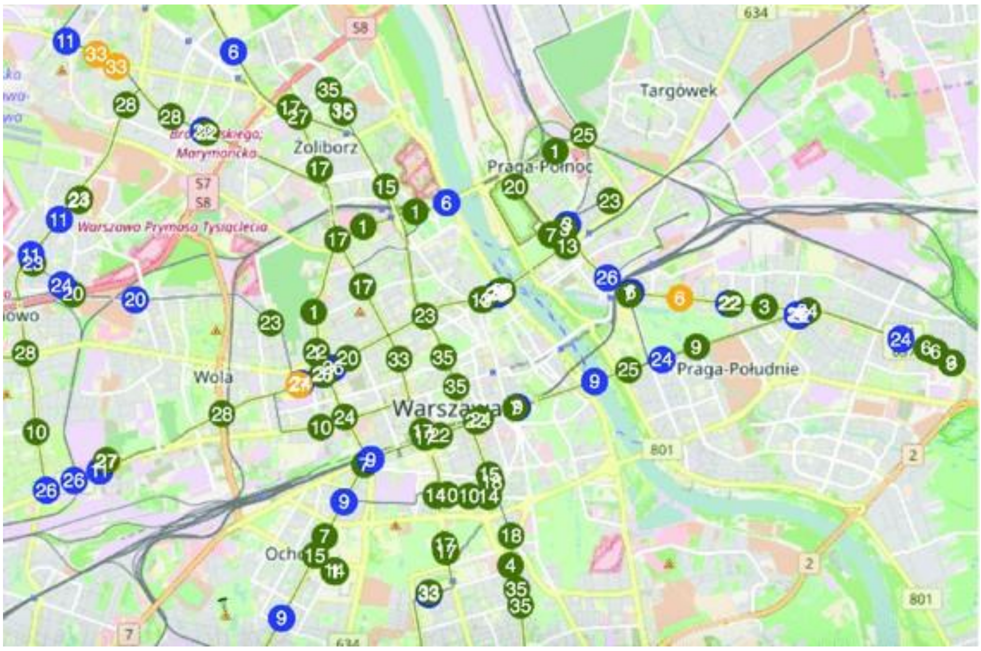


Fig. 32: Display of current delays of individual tram cars - numbers indicate the line, colour indicates the type of delay (blue = advance > 1 min., green = according to order, orange = delay > 3 min.).

Using data from February and March 2017, the results of the first experiment were evaluated in terms of analyzing the delay rate and the number of delayed tram cars. The published results of the analysis are shown in the following two aggregated graphs, broken down by day of the week and time of day.

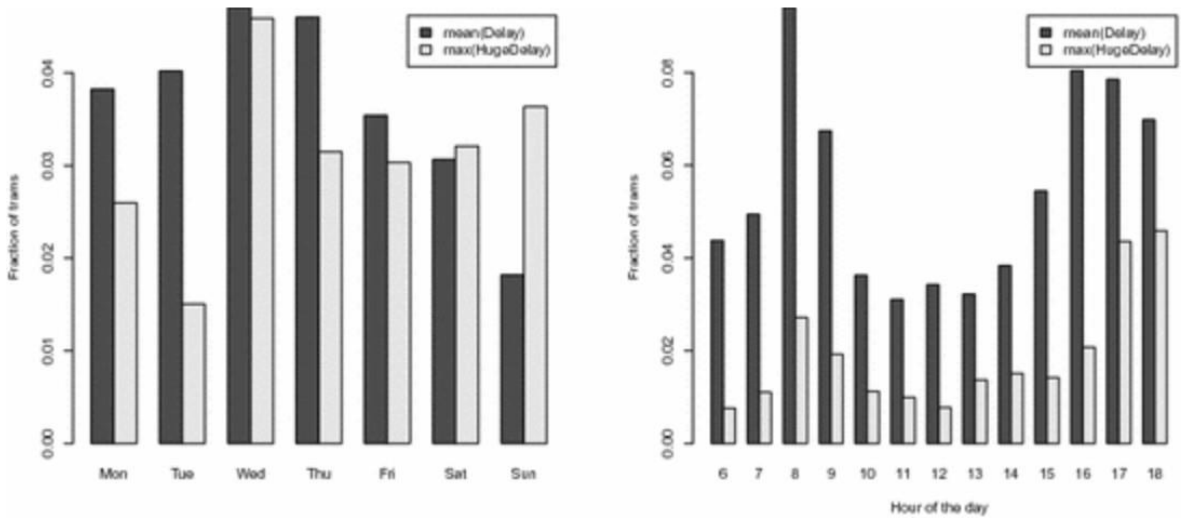


Fig. 33: Proportions of trams with average (dark bars) and significantly high (light bars) delays compared to the timetable, shown in a day-of-week (left) and time-of-day (right) breakdown.

4.4. Smart-city transport systems - an overview

In this chapter, for the completeness of the material of the technical part of the study, other relevant transport systems using modern technologies with a focus on artificial intelligence are listed, which can be used to determine especially the long-term plan for the development of the transport infrastructure of the IDS JMK and to a lesser extent also for the design of the current system for anomaly detection in terms of modularity and compatibility with future development. AI is used in many cities for a variety of transport-related applications, see for example [13] for an introduction, ranging from the monitoring of vehicle status, through the management of variable traffic infrastructure (traffic lights, tunnels, detour routes), estimating cyclist and pedestrian behaviour to fully autonomous public transport systems based on self-driving vehicles as part of the general Smart-city concept.

4.4.1. Autonomousbuses

Autonomous buses, i.e. self-driving cars without a driver or other, albeit remote, human operator, were originally introduced after the turn of the millennium in confined areas such as airports or large manufacturing plants. They are now also being introduced into mainstream public transport, particularly in China, Singapore and Finland.

Olli is the first autonomous bus to be implemented in a smart city infrastructure. It was introduced by Local Motors from the USA in 2016 and is currently being used in a number of locations for regular and conference or tourist transport (Chicago, Turin, Berlin, Yellowstone National Park). The production of the bus is largely based on 3D printing, and AI is used not only to monitor the surroundings and drive the car, but also to communicate with passengers about, for example, weather and tourist attractions.

Fig. 34: Olli autonomous bus equipped with artificial intelligence for driving.

Scania and Nobina is a combination of a truck and bus manufacturer and a public transport operator in the Nordic countries, who have launched a trial run of autonomous buses on regular routes in the Stockholm area in 2019. Two Scania Citywide LF electric buses connect the new residential area of Barkarby, which is approximately 20 kilometres from central Stockholm, with a nearby metro station. The buses are operating on a new special route with a length of 5 kilometres and four stops. It should be noted that Barkarby's public transport system has been described at expert level as the most modern in the world.

Fig. 35: ScaniaNobina autonomous bus equipped with AI for driving in the Barkarby area.

Bus parameters:

Model:	Scania Citywide LF (low-floor)
Vehicle length:	12 m
Drive:	Electric
Charging technology:	Depot charging
Carrying capacity:	80 passengers, of which 25 seats

4.4.2. Traffic management systems with AI

Surtrac is an intelligent traffic light control system from Rapid Flow Technologies that was initially installed in the City of Pittsburgh and then expanded to other cities. Surtrac's functionality is based on visual recognition of the current traffic situation, i.e., detecting vehicle, pedestrian and cyclist movements. Based on real-time detections, Surtrac creates a model of the current traffic and tracks where individual vehicles are heading. It then plans the efficient passage of vehicles through the intersection just by controlling the timing of traffic lights and the coordination of traffic flows between adjacent junctions. Testing of the Surtrac system in Pittsburgh has reduced total time spent in traffic by 25% and time spent waiting at green lights by 40%. Other announced improvements to the Surtrac system are also to use synchronization with vehicle navigation systems (Waze, Google Directions) to plan traffic flows based on vehicle destinations.

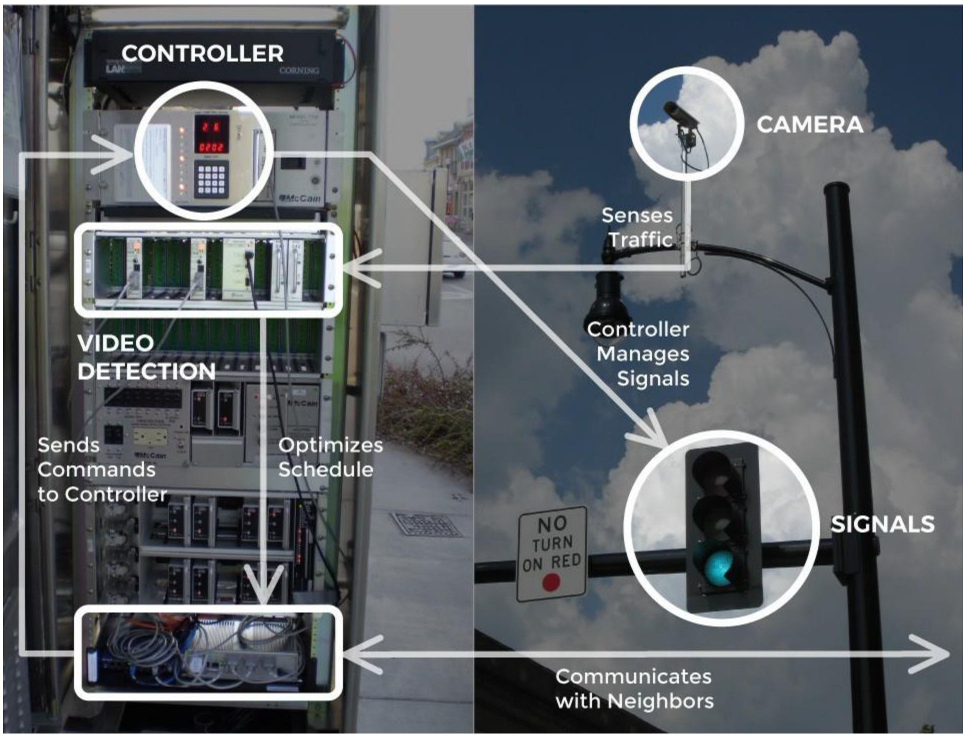


Fig. 36: Surtrac's technical means for AI-assisted traffic management.

Siemens Mobility is a well-known term for the monitoring system originally implemented in the city of Bengalur, India, which adjusts the timing of traffic lights according to the detection of the current traffic density.

City Brain is an Alibaba Cloud project introduced in 2017. Like the previous systems mentioned, it optimises traffic flows by controlling traffic lights using data from CCTV cameras. In 2019, 23 cities have already used this system. A pilot implementation of the comprehensive system was carried out in the Chinese city of Hangzhou (approx. 2 million inhabitants) with the following results:

- traffic time on the 22 km long motorway section around the city was reduced by an average of 4.6 minutes,
- it takes an average of 2.6 seconds to pay for parking,

- the arrival time of the emergency services at the accident scene was reduced by 7 min. on average.

Seoul is the capital of South Korea with a population of about 10 million, where AI in the form of Big Data was used in 2014 to analyse and design night bus routes (night-owls). The location of residents provided by the mobile operator and the destination of taxi vehicles during the desired night hours (midnight to 5 a.m.) were used to design the bus routes. Data collection was carried out over a period of one month. In the first phase, the route scheme was designed in a conventional planning manner based on available public transport information. In the second phase, a geographic visualization of the population density at a given nighttime hour was created based on mobile operator data and taxi data, and the design of the routes was optimized. The resulting design was positively evaluated by Seoul residents who previously lacked night bus routes.

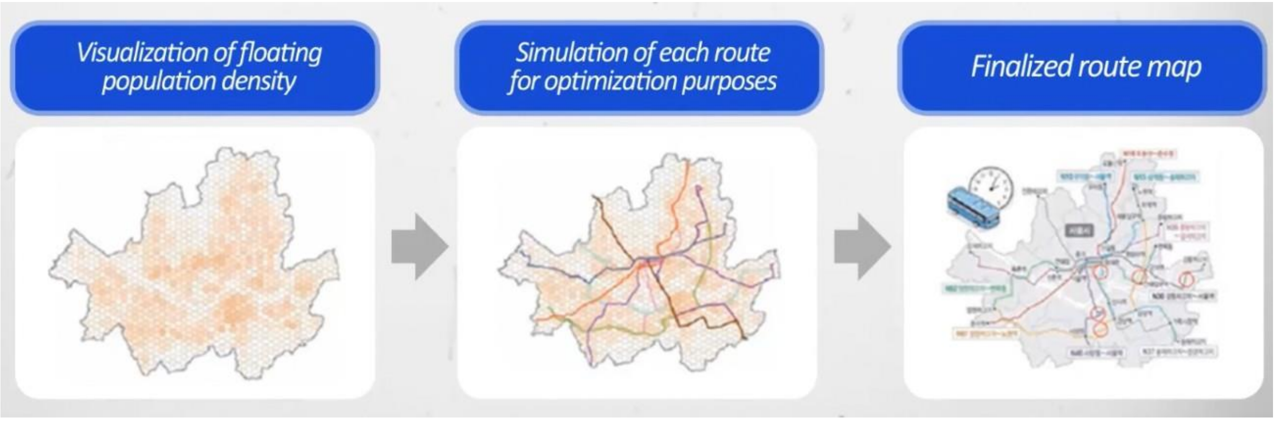
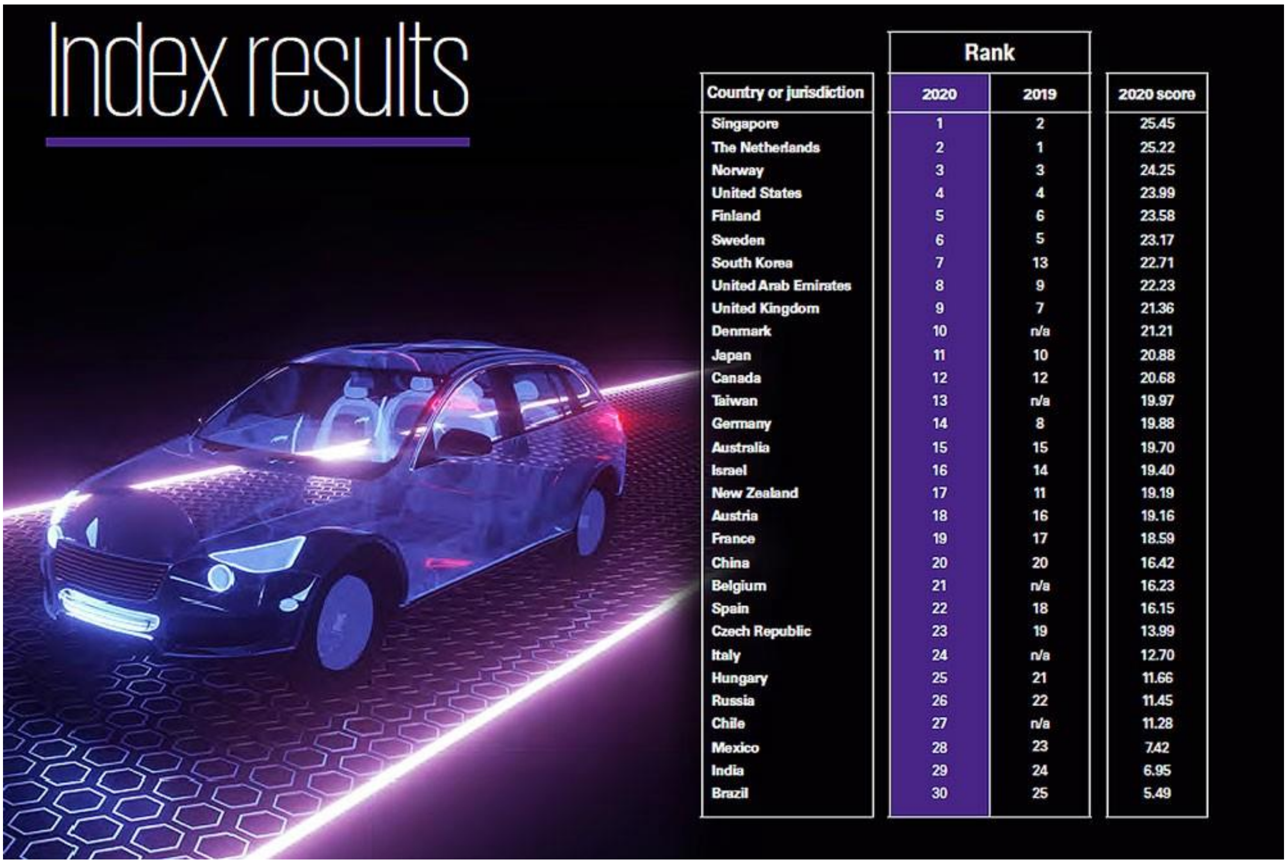


Fig. 37: Design of night lines in Seoul.

As a highly urbanised city-state with the third highest population density in the world (over 7,000 inhabitants per km^2), **Singapore** has been engaged in AI-based traffic management since 1998. In that year, Singapore introduced Electronic Road Pricing (ERP Singapore) for all types of road vehicles, which eliminates the need for vehicles to stop at toll gates or slow down at checkpoints when paying road tolls. The system uses a network of sensors and CCTV cameras to monitor traffic on the road as well as extract vehicle registration plates. Each vehicle is equipped with a so-called In-vehicle Unit unit used for immediate payment of tolls. An upgraded version is the Next Generation ERP Singapore, in which all vehicles are equipped with GPS locators whose data (location, speed, time) is collected by internal systems. Thus, in real time, the control system will notify the vehicle driver of information about an impending congestion or other road closure and suggest an alternative route. At the same time, the NG ERP Singapore system is used for parking payments, with charges aggregated automatically according to GPS data. The City of Singapore currently

Strobo's autonomous buses are currently being tested and are scheduled to go live in 2022. In terms of an example, it is worth noting that the city-state of Singapore is ranked first in KPMG's Autonomous Vehicles Readiness Index, while the Czech Republic has fallen to 23rd place from the previous 19.



Country or jurisdiction	Rank 2020	Rank 2019	2020 score
Singapore	1	2	25.45
The Netherlands	2	1	25.22
Norway	3	3	24.25
United States	4	4	23.99
Finland	5	6	23.58
Sweden	6	5	23.17
South Korea	7	13	22.71
United Arab Emirates	8	9	22.23
United Kingdom	9	7	21.36
Denmark	10	n/a	21.21
Japan	11	10	20.88
Canada	12	12	20.68
Taiwan	13	n/a	19.97
Germany	14	8	19.88
Australia	15	15	19.70
Israel	16	14	19.40
New Zealand	17	11	19.19
Austria	18	16	19.16
France	19	17	18.59
China	20	20	16.42
Belgium	21	n/a	16.23
Spain	22	18	16.15
Czech Republic	23	19	13.99
Italy	24	n/a	12.70
Hungary	25	21	11.66
Russia	26	22	11.45
Chile	27	n/a	11.28
Mexico	28	23	7.42
India	29	24	6.95
Brazil	30	25	5.49

Figure 38: KPMG's State Readiness Index for Autonomous Transport.

With the introduction of technologies using AI, Singapore has become a leader in solving traffic congestion, with the average speed of traffic on the city's main roads being 27 km/h, while in London, for example, the same parameter is 16 km/h.

5. TECHNICAL MEANS AND KNOWLEDGE REQUIREMENTS

As a counterbalance to the often unique solution to a previously unaddressed problem, the use of modern AI elements also brings with it higher requirements for technical resources and the need for increased operator expertise. Considering the aforementioned technology boom in 2009, when nVidia's special computing cards in the form of Graphics Processing Units (GPUs) were launched, it is easy to deduce that the market in this area is still very dynamic and new technologies appear regularly every year. Conventional office computers are not suitable for tasks using AI architectures because they lack the capability for massively parallelised computations that are typical for AI tasks. Indeed, AI architectures typically contain a high number of repetitive computational blocks of common mathematical operations, such as convolution in the case of neural networks.

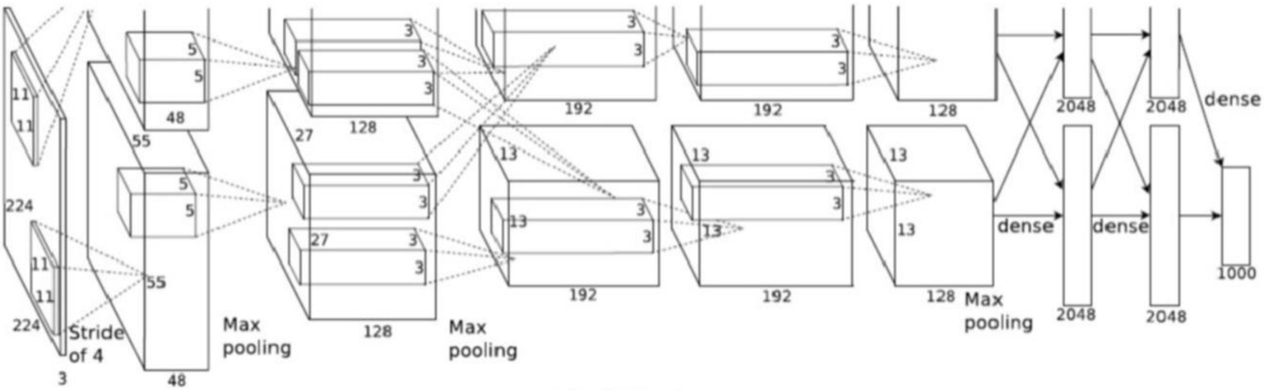


Fig. 39: Example of the internal architecture of the image information processing part of a convolutional neural network.

The architecture of a personal computer processor (CPU - Central Processing Unit) is based on a large cache that allows serial processing of advanced logic operations in a few arithmetic-logic units (ALU) (modern processors typically contain 4 or 8 so-called cores). It is therefore essentially the opposite of parallelization. For this reason, GPU architectures, or GPGPUs (General Purpose GPUs), whose architecture is specifically designed to be massively parallelized, are applied for the needs of not only AI tasks. The very first GPGPUs contained orders of magnitude more computing cores than the original CPUs, and modern GPGPU cards even contain thousands of computing cores. Recently, especially with the development of FW, the long known Field Programmable Gate Array (FPGA) technology has also come to the forefront of AI task implementation, which until now required very high requirements for knowledge of the specific HW architecture and was thus incompatible with the predominantly programming nature of AI task implementation.

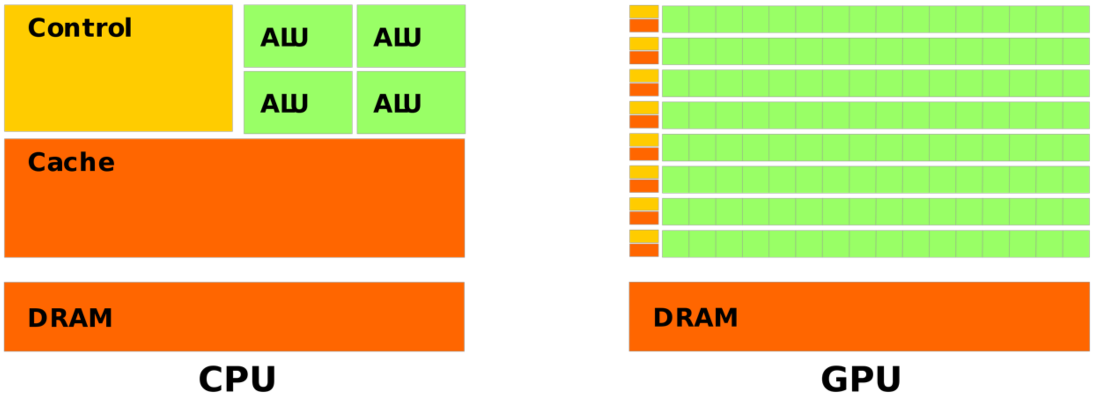


Fig. 40: CPU vs. GPU architecture - an illustration of the difference in the number of computing cores.

To support parallel computing, GPGPUs from different vendors are designed in a specific proprietary structure. In the case of nVidia, for example, this is the well-known CUDA (Compute Unified Device Architecture) technology, which enables efficient execution of programs written in supported programming languages (here specifically C/C++, Fortran, etc.).

Fig. 41: NVIDIA Quadro M5000 GPGPU with 2048 cores (MC approx. 40 thousand CZK).

It is possible to use a commonly available workstation to work with the above mentioned HW, but it is necessary to size it for the HW, especially the power supply, operating memory and storage. The cost of a very basic HW workstation enabling the development and partly also the subsequent operation of the AI mechanism should thus be calculated at a minimum of 50 thousand CZK. CZK (in 2021).

In addition to the above technologies suitable for solving tasks using AI methods, there is another specific architecture called TPU (Tensor Processing Unit). These special tensor computing units are used for demanding tasks mainly in science and research, or for extremely data-intensive industrial tasks.

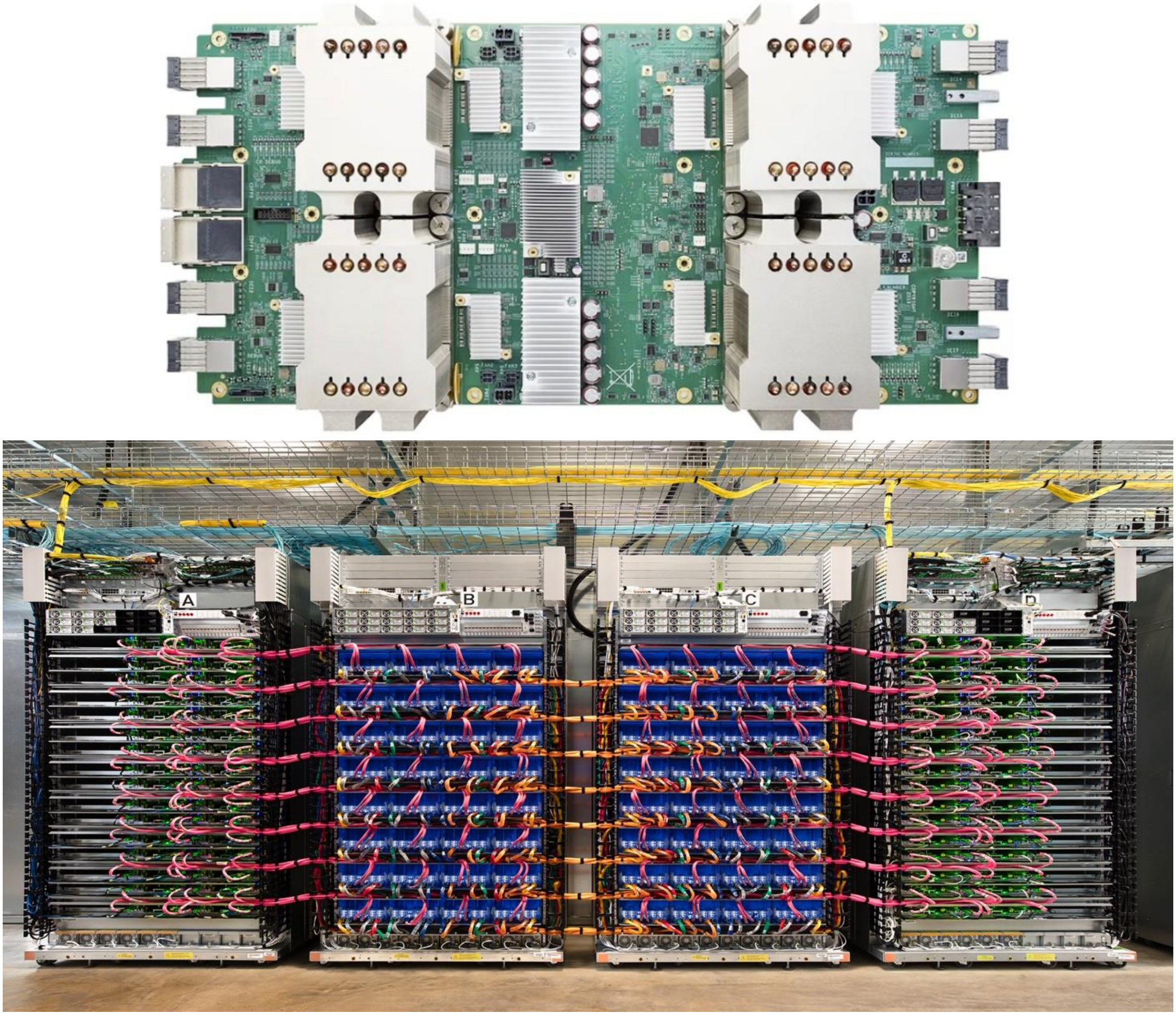


Figure 42: Example Google Cloud TPU (top) and Google's high-end cloud array with 64 TPUs and a total compute power of 11.5 PFLOPS.

From the point of view of working with the above mentioned HW, in addition to the usual programming knowledge, knowledge of AI methods and, above all, knowledge of special libraries designed for the implementation of AI architectures is required. These are constantly evolving and currently various combinations of development environments and libraries are often used. Among the most well-known can be mentioned Tensorflow, Pytorch, Keras, etc. using C/C++, Python, CUDA and other programming languages.

Regarding the design of a system for anomaly detection in public transport in the IDS JMK, it is necessary to consider at least for the design and implementation a professional workplace with experience in designing AI architectures in transport or industry. For the subsequent operation of the system, no special knowledge of the operator is required, but it is necessary that the dispatcher operator has at least been trained with basic information about the specific way the system works.

6. LIST OF ABBREVIATIONS

AIArtificial Intelligence / Artificial Intelligence

ANNArtificial Neural Network / Artificial Neural Network

AVRIAutonomous Vehicles Readiness Index

BeiDouGlobal Satellite Positioning System (China)

CNNConvolutional Neural Network / Convolutional Neural Network

CPUCentral Processing Unit / Central Processing Unit

FPGAField Programmable Gate Array GalileoGlobal Positioning Satellite System (EU ESA)

GLONASS Global Navigation Satellite System / Global Positioning Satellite System (Russia)

GLCM Gray Level Cooccurrence Matrix (matrix used in texture analysis) GPS Global Positioning System / Global Positioning Satellite System (USA) GPU Graphic Processing Unit

GPGPUGeneral Purpose GPU / General Purpose Graphical Computing Unit

GTFSGeneral Transit Feed Specification (Public Transport Schedule Format)

IDS JMKIntegrated Transport System of the South Moravian Region LSTMLong Short-term Memory (recurrent ANN type)

MLMachine Learning / Machine Learning

MLP-RLMultilayer Perceptor - Linear Regression / Multilayer perceptron (artificial neural network architecture)

R-CNNRegion-based Convolutional Neural Network (deep learning architecture)

SSDSingle Shot Detector (deep learning architecture)

TPUTensor Processing Unit / Tensor Computing Unit YOLOYou Only Look Once (Deep Learning Architecture)

7. REFERENCE

List of literature used:

[1] RUSSELL, S. J., & NORVIG, P. Artificial intelligence a modern approach. Boston, Pearson, 2020. ISBN 9780134610993.

[2] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning: with Applications in R. New York: Springer, 2021. ISBN 978-1071614174.

[3] Abduljabbar, R:, Dia, H., Liyanage, S., Bagloee, S. A. Applications of Artificial Intelligence in Transport: An Overview. MDPI Sustainability, vol. 11. 2019. ISSN 2071-1050.

[4] Transportation Research Board: Artificial Intelligence Applications to Critical Transportation Issues. Circular E-C168. Washington, DC, 2012. ISSN 097-8515.

[5] The European Economic and Social Committee & European Committee of the Regions. Sustainable and Smart Mobility Strategy - putting European transport on track for the future. European Commision, COM(2020) 789 final, Brussels. PermaLink: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0789.

[6] Mikolaj L., Remigiusz B., Lukaszand S., Rychhlik M., Slusarczyk P. Public Transport Vehicle Detection Based on Visual Information. 2004. Communications in Computer and Information Science, vol. 429, ISBN 978-3-319- 07569-3 2.

[7] Impedovo, D.; Balducci, F.; Dentamaro, V.; Pirlo, G. Vehicular Traffic Congestion Classification by Visual Features and Deep Learning Approaches: A Comparison, Sensors 2019, 19, 5213. https://doi.org/10.3390/s19235213.

[8] Li Wei, Dai Hong-ying, Real-time Road Congestion Detection Based on Image Texture Analysis, Procedia Engineering, Volume 137, 2016, Pages 196-201, ISSN 1877-7058, https://doi.org/10.1016/j.proeng.2016.01.250.

[9] Linsey Xiaolin Pang, Sanjay Chawla, Wei Liu, Yu Zheng, On detection of emerging anomalous traffic patterns using GPS data, Data & Knowledge Engineering, Volume 87, 2013, Pages357-373 , ISSN0169-023X , https://doi.org/10.1016/j.datak.2013.05.002.

[10] A. Tsiligkaridis and I. C. Paschalidis, Anomaly detection in transportation networks using machine learning techniques, 2017 IEEE MIT Un- dergraduate Research Technology Conference (URTC), 2017, pp. 1-4, doi: 10.1109/URTC.2017.8284194.

[11] Xiaocai Zhang, Yi Zheng, Zhixun Zhao, Yuansheng Liu, Michael Blu- menstein, Jinyan Li, Deep learning detection of anomalous patterns from bus trajectories for traffic

insight analysis, Knowledge-Based Systems, Volume 217, 2021, 106833, ISSN 0950-7051.

[12] Luckner, Marcin & Karwowski, Jan. (2017). Estimation of Delays for Individual Trams to Monitor Issues in Public Transport Infrastructure. 518-527.

[13] Lopez Conde, Maria; Twinn, Ian. 2019. How Artificial Intelligence is Making Transport Safer, Cleaner, More Reliable and Efficient in Emer- ging Markets. EMCompass;Note 75. International Finance Corporation, Washington, DC. © International Finance Corporation. https://openknowledge.worldbank.org/handle/10986/33387.

[14] BIN OTHMAN, Muhammad Shalihin and Gary TAN. Machine Learning Aided Simulation of Public Transport Utilization. 2018 IEEE/ACM 22nd International Symposium on Distributed Simulation and Real Time Applications (DS-RT). IEEE, 2018, 2018, 1-2. ISBN 978-1-5386-5048-6. Available from: doi:10.1109/DISTRA.2018.8601011.

[15] DEB, Bilash, Salehin Rahman KHAN, Khandker TANVIR HASAN, Ashikul Haque KHAN and Md. Ashraful ALAM. Travel Time Prediction using Machine Learning and Weather Impact on Traffic Conditions. 2019 IEEE 5th International Conference for Convergence in Technology (I2CT). IEEE, 2019, 2019, 1-8. ISBN 978-1-5386-8075-9. Available from: doi:10.1109/I2CT45611.2019.9033922.

[16] HUANG, Yanjun, Haiying LI, Xi JIANG, Qi SUN and Chengcheng SU. A Machine Learning Based Method for Short-term Urban Rail Transit Passenger Flow Prediction. 2020 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC). IEEE, 2020, 2020-8-5, 335-339. ISBN 978-1-7281-7050-3. Available from: doi:10.1109/SDPC49476.2020.9353156.

[17] J. Svanberg. Anomaly detection for non-recurring traffic congestions using long short-term memory networks (LSTMs), 2018.

[18] G. Qin, Z. Huang, Y. Xiang, and J. Sun. Probdetect: A choice probability-based taxi trip ano- maly detection model considering traffic variability. Transportation Research Part C: Emerging Technologies, 98:221-238, 2019.

[19] F. Sun, A. Dubey, and J. White. Dxnat - deep neural networks for explaining non-recurrent traffic congestion. 2018.

[20] X. Cai. Collaborative prediction for bus arrival time based on cps. Journal of Central South University, 21(3):1242-1248, 2014.

[21] The 2030 Agenda for Sustainable Development, United Nations publication, Santiago, 2018. ISBN 978-92-1-122011-7.

[22] Krásenský, D., Sklenář, M. Public transport of the South Moravian Region: ten years of system and technological integration of IDS of the South Moravian Region. ČD Scientific and Technical Proceedings No. 34/2012.

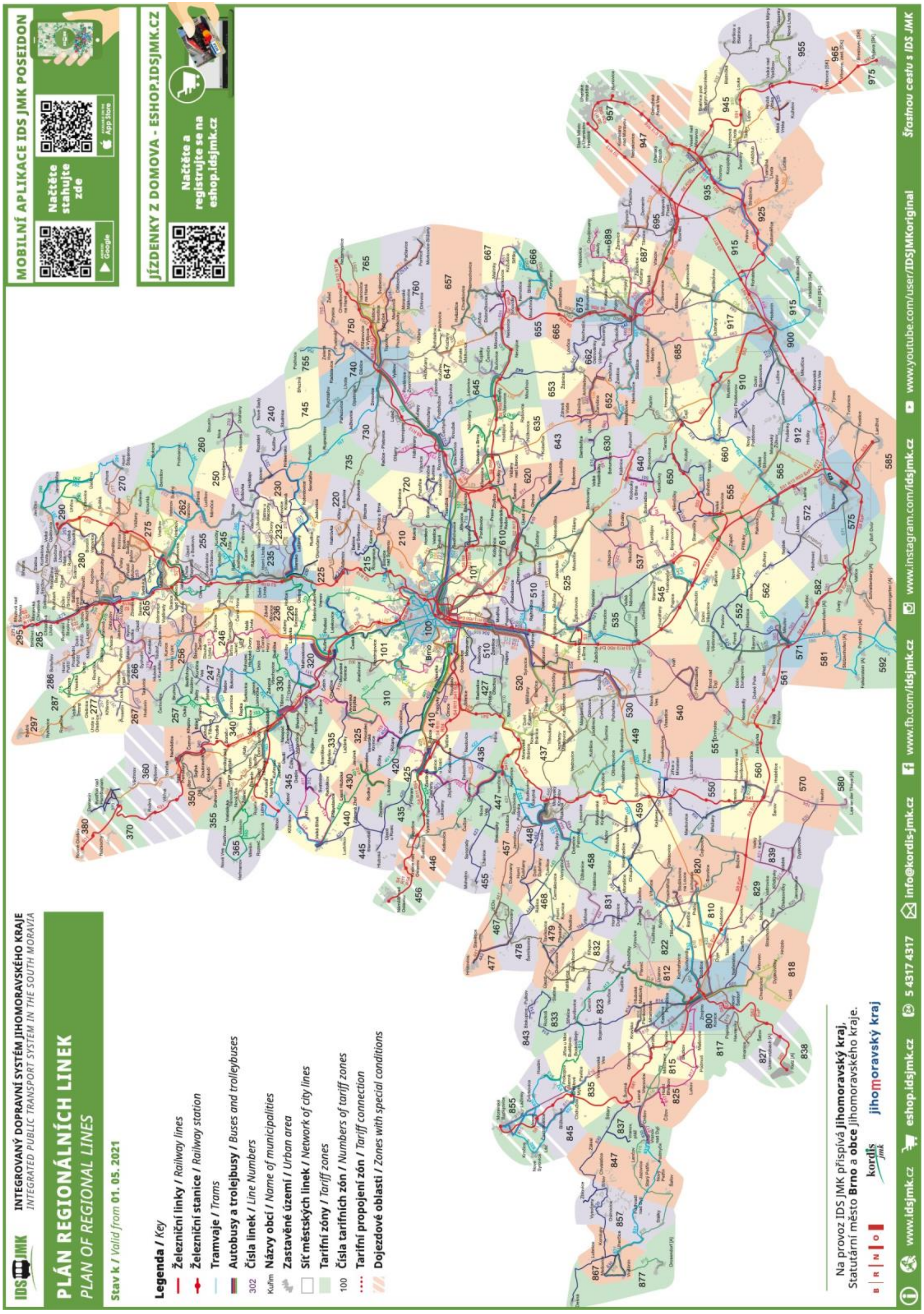
[23] Kordis JMK, Annual Report 2019. Online, available from http://www.idsjmk.cz.

[24] Cruz, M., Barbosa, L. Learning GPS Point Representations to Detect Anomalous Bus Trajectories. IEEE Access - Open Access journal, vol. 8, 2020. DOI 10.1109/ACCESS.2020.3046912.

[25] Fangzhou, S. Algorithms for Context-Sensitive Prediction, Optimization and Anomaly Detection in Urban Mobility. Dissertation thesis, 2018, Faculty of the Graduate School, Vanderbilt University, Tennessee.

8. *ANNEXES*

8.1. Plan of regional lines of IDS JMK



8.2. Network of daily transport lines of the city of Brno

8.3. IDS JMK statistics

Počet linek v IDS JMK	322
Počet vlakových linek	24
Počet tramvajových linek v Brně (čísla 1 až 12)	11
Počet trolejbusových linek v Brně (čísla 20 až 40)	13
Počet autobusových linek v Brně (čísla 40 do 100)	47
Počet linek městské dopravy v Adamově	1
Počet linek městské dopravy v Blansku	4
Počet linek městské dopravy v Břeclavi	9
Počet linek městské dopravy v Bystřici n/P	1
Počet linek městské dopravy v Hodoníně	4
Počet linek městské dopravy v Kyjově	3
Počet linek městské dopravy ve Vyškově	4
Počet linek městské dopravy ve Znojmě	7
Počet regionálních autobusových linek	194
Počet zón	161

8.4. IDS JMK carriers

V IDS JMK je začleněno celkem 24 dopravců.

Železniční dopravu zajišťují:

- České dráhy, a.s.
- Železničná spoločnosť Slovensko, a.s.
- Regiojet a.s.
- Arriva vlaky, s.r.o.
- ÖBB

Dopravu na území města Brna zajišťuje především:

- Dopravní podnik města Brna, a.s.

Dopravu na regionálních autobusových linkách zajišťují:

- ADOSA a.s.
- ARRIVA MORAVA a.s.
- DOPAZ s.r.o.
- BDS-BUS
- BORS Břeclav a.s.
- BORS BUS s.r.o.
- ČAD Blansko a.s.
- ČSAD Brno holding, a.s.
- ČSAD Hodonín a.s.
- ČSAD Kyjov Bus a.s.
- ČSAD Tišnov, spol. s r.o.
- Dopravní podnik města Brna
- FTL - First Transport Lines
- ICOM transport a.s.
- Tourbus, a.s.
- TRADO-BUS, s.r.o.
- VYDOS BUS a.s.
- ZDAR, a.s.
- Zlatovánek spol.s r.o.
- Znojemská dopravní společnost - PSOTA, s.r.o.

8.5. GTFS specification data files

Filename	Required	Defines
agency.txt	Required	Transit agencies with service represented in this dataset.
stops.txt	Required	Stops where vehicles pick up or drop off riders. Also defines stations and station entrances.
routes.txt	Required	Transit routes. A route is a group of trips that are displayed to riders as a single service.
trips.txt	Required	Trips for each route. A trip is a sequence of two or more stops that occur during a specific time period.
stop_times.txt	Required	Times that a vehicle arrives at and departs from stops for each trip.
calendar.txt	Conditionally required	Service dates specified using a weekly schedule with start and end dates. This file is required unless all dates of service are defined in calendar_dates.txt.
calendar_dates.txt	Conditionally required	Exceptions for the services defined in the calendar.txt. If calendar.txt is omitted, then calendar_dates.txt is required and must contain all dates of service.
fare_attributes.txt	Optional	Fare information for a transit agency's routes.
fare_rules.txt	Optional	Rules to apply fares for itineraries.
shapes.txt	Optional	Rules for mapping vehicle travel paths, sometimes referred to as route alignments.
frequencies.txt	Optional	Headway (time between trips) for headway-based service or a compressed representation of fixed-schedule service.
transfers.txt	Optional	Rules for making connections at transfer points between routes.
pathways.txt	Optional	Pathways linking together locations within stations.
levels.txt	Optional	Levels within stations.
feed_info.txt	Conditionally required	Dataset metadata, including publisher, version, and expiration information.
translations.txt	Optional	Translated information of a transit agency.
attributions.txt	Optional	Specifies the attributions that are applied to the dataset.