

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Risk Assessment Inference Approach based on Geographical Danger Points using Student Survey Data for Safe Routes to School

WENQUAN JIN¹, AZIMBEK KHUDOYBERDIEV² and DOHYEUN KIM²

¹Big Data Research Center, Jeju National University, Jeju 63243, Republic of Korea; wenquan.jin@jejunu.ac.kr (W.J.)

²Department of Computer Engineering, Jeju National University, Jeju 63243, Republic of Korea; azimbekkhudoyberdiev@jejunu.ac.kr (A.K.); kimdh@jejunu.ac.kr (D.K.)

Corresponding author: Dohyeun Kim (e-mail: kimdh@jejunu.ac.kr).

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (NRF-2019R111A1A01062456), and this research was supported by Energy Cloud RD Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT (2019M3F2A1073387). Any correspondence related to this paper should be addressed to Dohyeun Kim.

ABSTRACT Safe Routes to School is very important for students to have good physical and psychologically healthy in school life. For providing safe routes based on risk analysis, finding out dangerous points and areas can be a target to avoid dangerous locations by pedestrians and drivers. However, analyzing the risk assessment to derive the safe routes requires a large amount of data with a certain time of observation by experts. Deep learning is a solution to provide information regarding safe routes based on expert knowledge. In this paper, we propose a risk assessment inference approach using a Recurrent Neural Network (RNN) model with Long-Short Term Memory (LSTM) cells based on geographical information for safe routes to school. However, geographical information including coordinates is difficult used in learning-based inference models because of the series of float values. For training the RNN model with the geographical data, coordinates of routes and danger points are translated to be geohash through the geohash converter. The geohash data with other data of features are fused and inputted to the one-hot encoder. The one-hot encoded data is used in the inputs of the RNN model to train the LSTMs. The input data of the training model is derived by the risk index model that is proposed to calculate the risk index based on distances of route coordinates and danger points. Therefore, the risk index is correlated with the training dataset. Through the proposed inference approach, the geographical information including multiple coordinates is enabled to be trained by RNN as a geohash-based input string. Moreover, the input string with other features is fused to support the one-hot encoding to get a better result in RNN models.

INDEX TERMS Recurrent Neural Network, Long-Short Term Memory, One-Hot Encoding, Geographic Information System, Geohash, Risk Assessment, Safe Routes to School

I. INTRODUCTION

ROAD traffic accidents are one of the wide-ranging and most crucial problems for humanity in the world. World health organization statistics describe that road crashes are the second main cause of death globally among young people aged from 5 to 29. Every year more than 1.2 million people are killed by road crashes and nearly 50 million people are injured [1]. Road accidents or dangers generally result of five different factors including human (e.g., driver or pedes-

trian behavior), environmental (e.g., condition of weather), infrastructure (e.g. road design), traffic condition (e.g. traffic congestion) vehicle-related factors (e.g. age or size of the vehicle) [2]. Over the last decades, remarkable practical and methodological developments have been achieved for decreasing dangers and accident rates on roads based on heterogeneous technologies such as stochastic, heuristic and fuzzy [3]–[6]. However, the lack of knowledge in terms of human factors, driving issues, vehicle mechanisms are the

main problem for solving human daily traveling accidents as a whole [7]. Today many researchers are paying attention to improve vehicle design and control functionalities to decrease risk levels, as well as, various road assessment programs are promoting redesign roads to minimize dangers on the roads [8]. Vehicles are believed as the main factor of danger on roads since they are the first actors in the accidents. For these reasons, researchers have been trying to suggest intelligent solutions that can reduce the risk of dangers on the ways based on historical or current sensing data in a specific road zone [9]–[11]. The role of the intelligent services is to sense the environment and analyze the data regarding the current zone, then predict and detect future risks.

Safe Routes to School (SRTS) enables walking and bicycling to school safer which enhance children's health and well-being, reduce traffic congestion near the school and improve community life quality [12]. However, the complexity of the roads makes many parents drive their children to avoid potential dangers [13]. Most of the studies found that road infrastructure also has an essential impact on the risk of traveling both by vehicles and by walk [14]–[16]. In particular, there are intersections on roads which are generally known as black spots for school children as well as for all road users [17]. Black spots are specific locations where the risk of having accidents and expectations for potential conflicts are higher than the rest of the roads [18], [19]. The risk and the number of dangers on routes are generally led by traffic conditions such as the composition of traffic, vehicle flows, or volumes. During peak hours, the danger level and the number of accidents increases, its influence on not only vehicle drivers but also pedestrians [20]–[22]. Based on these data, Geographic Information System (GIS) provides capturing, storing, manipulating, analyzing and management functionalities [23], [24]. The geographic information can keep track of not only things, activities, and events, but also where these things, activities, and events exist or happen [25], [26]. Therefore, sufficient data enables routes to be safer by removing dangers based on predictable parameters.

Recently, Machine Learning (ML) algorithms have been receiving extensive attention to determine future occurrences or recommend actions to achieve optimal outcomes in various fields such as healthcare, finance, and transportation to prevent the risks [27]–[30]. Deep learning algorithms have become one of the hottest technologies in many research fields. Especially, Recurrent Neural Network (RNN) with Long-Short Term Memory (LSTM) has been applied successfully in sequential data such as time series forecasting, natural language translation and speech recognition [31]–[35]. LSTM-based RNN trains data by sequential processing over time. However, the deep learning model can be easily corrupted for learning small values because of vanishing gradients [36], [37]. The geographical information is based on coordinates that are comprised of a couple of float data. Multiple coordinates are difficult to be fused with other features for a one-hot encoded dataset to train an LSTM-based RNN model [38]–[40]. However, predicting the risk

of routes requires a set of coordinates that illustrates the information of a path of route and danger points. Therefore, the LSTM-based RNN model is difficult to train a set of coordinates for predicting the risk assessment using geographical information.

For awareness of dangers to enable SRTS, in this paper, we propose a risk inference approach model that uses geographical information including a path of route and danger points to predict the route risk index. For predicting the risk index, an LSTM-based RNN model is proposed to use one-hot encoded data that involves information of coordinates of routes and danger points. For combining the coordinates with other strings, the coordinates are translated to a set of geohash [41], [42]. The geohash data is a widely used standard for describing the location using a short alphanumeric string, more precisely, location is described with latitude and longitude values. The transformed geohash data and other data features including danger types, transportation, and gender are combined and used as an input parameter to the one-hot encoder. For preparing risk index data to be used in the RNN training model, we propose a risk index model based on calculating the distance of danger points and routes using a mathematical equation. For the implementation and experiment, the RNN models are configured to be 100, 300, 500 times training epochs to test the performance of the proposed inference approach using the geographical information. Furthermore, we collect the predicted risk index from the proposed inference model to compare with the risk index of the equation-based risk index model for experimenting with the proposed risk inference approach model using data of school routes.

The rest of the paper is structured as follows. Section II introduces the related works regarding risk assessment approaches based on geographical information and inference systems. Section III introduces the proposed risk inference approach through the architecture of models to depict the data pipelines and specifications. Section IV introduces data processing and presentation to present the data that is used in the proposed inference approach. Section V presents experimental results including the predicted risk index and danger level. Section VI presents the performance evaluation through comparisons of risk indexes and danger levels. Finally, we conclude this paper and introduce our future directions in Section VI.

II. RELATED WORKS

In recent years, several strategies have been developed to decrease traffic problems in the world, many metropolitan organizations have started to promote the development and redesigning of vehicles and roads structure to decrease the risk of dangers on roads [43]. Reducing the risk of dangers offer considerable benefits to society in terms of personal safety, health improvement, better air quality, and fewer worries [44]. For monitoring and controlling the risk of dangers on the roads, GIS and ML algorithms have been using widely. GIS has been achieving a huge reputation, which can

provide better visualization of the large dataset for decision making and analyzing process. GIS-based maps help for finding out the crash spots, risky points and danger zones in motorways and highways [45]. The road safety analysis in M-25 motorway, the GIS supports the required data of traffic, characteristics of the road, and accidents with 70 segments [46]. High risky crash zones have been described clearly in Shanghai expressways using GIS-based applications [47]. In Tehran and Belgium, GIS-based applications have been successfully applied to map the black zones to increase the attention of drivers and pedestrians in such zones [48], [49]. The development of GIS-based applications has helped to model crash data in a short-range, as well as based on short-range identification long-range crash area points clarified continuously using GIS in Belgium [50].

The recent improvements in the efficiency of Remote Sensing (RS) and GIS technologies have initiated a revolution in hydrology, particularly in flood management, which can fulfill all the requirements for flood prediction, preparation, prevention, and damage assessment [51]. Among different GIS-based flood models presented in the literature, artificial neural networks [52], frequency ratio [53], logistic regression [54], adaptive network-based fuzzy inference system [55], multi-layered feed-forward network [56], decision trees [57], [58], and support vector machines [59], [60] are the most widespread techniques that utilize RS and GIS tools [61]. Although flood forecasting and prediction models are available, the accuracy of flood prediction maps remains a critical issue. In flood modeling, high accuracy for flood prediction mapping should be achieved, and thus, new and efficient models should be explored to increase the accuracy. Flood risk can be expressed as a combination of hazard and vulnerability [62]–[64]. In particular, the risk is a mathematical expectation of the vulnerability (consequence) function. Flood probabilities are determined to produce flood hazard maps. Hydraulic models may result in uncertainties because they require complete and sufficient hydrological data [65], [66]. Therefore, using RS data and GIS-based models can be considered a complementary approach to flood modeling.

Khan et al. suggested an approach based on Hidden Markov Model (HMM) for destination and route prediction [67]. They have suggested a lightweight algorithm for predicting driver routes and destinations. The last visited road links were used as an input parameter for the HMM-based prediction algorithm, a client application supports real-time visual predicted risks with GPS. Vehicle to the cloud, cloud to vehicle connectivity-based road and route risk assessment planning have been suggested in [68]. They have developed a crash prediction model using Artificial Neural Network (ANN) and more than 30,000 road segments and 144,821 crashes data which was collected from the highway. Their proposed system analyzes how current weather, current time and day of the week influence on safe route planning. Zheng et al. have proposed a road traffic risk assessment framework based on HMM [69]. They analyzed the steering angle and the velocity of the vehicle to predict the movement of the

vehicle over a period, also they can calculate the expected path of the neighbor car. The HMM algorithm is used for calculating the steering angle of each vehicle, as well as, road safety level also considered as an input parameter to train the HMM algorithm. They have successfully tested their proposed framework on real-time and road traffic risks can be monitored continuously based on their algorithm. In this study, only the motion of vehicles is mentioned, but in the real road includes various types of participants such as cyclists, walkers, animals, and pedestrians.

A study is presented in [70] that provides GIS and ANN-based road safety risk evaluation research. The authors selected two motorways in Belgium (E-313 and E-314), and they divided the lengths of the motorways to 67 segments, the crash data used in this research includes occurred crash location from 2010 and 2012. Vehicle position, horizontal and vertical crash points, speed, environment condition, and flow are taken as an input parameter for training the ANN model. The output of the prediction model describes the riskiest points on a motorway, as well as, the length of the riskiest segments. The predictive capability of GIS has been improved by using ANN algorithms in a wide range of applications [71]. The combination of GIS and ANNs are a popular solution in agriculture, meteorology, geoscience, and land irrigation. GIS-based RNN model is being used to predict the injury points and crashes in Malaysia expressways [72]. Some studies have been focusing on improvements in vehicle mechanisms, while others are paying attention to the design of the roads to decrease risk levels in routes. The routing security and predicting the safest routine based on risk factors concept has been presented in [73]. They have developed a cost-effective infrastructure, their proposed application operates in three steps, firstly based on the starting point and destination the system collects the list of risks from the database. Secondly, the data is preprocessed for removing stop words from the original data. Finally, the system suggests the most optimal and closest route to users. According to the comparisons of google safe route suggestions, their proposed system more clear and safer. However, the authors did not discuss predictive algorithms on how they configured and which type of algorithm is used for the system.

Several ML approaches are proposed to learn sequences for deriving the results such as classes, floats, identifies and sequences in various studies. Guo et al. [74] proposed a sequential classifier based on Support Vector Machine (SVM) that reads the image data as sequences to classify the multitemporal remote sensing images. Meng et al. [75] proposed an SVM-based protein identification solution that extracts the information from the training dataset and delivers to the learning for reducing the complexity of data. Zahidul et al. [76] proposed a text classification approach based on the random forest that is suitable to process the high dimensional noisy data through extracting features using trees. Moreover, for classifying the text data, Xu et al. [77] proposed an improved random forest that is developed with a novel weighting and tree selection methods to reduce subspace size and

improve performance. Sutskever et al. [34] proposed a multi-layer LSTM in the RNN for translating English to French which achieves outstanding accuracy and brings the approach into notice [78]. The use of Evolutionary Computing (EC) for machine learning is presented from various perspectives such as feature selection and classification, regression and deep learning [79]. In deep learning the EC provides an optimal solution to machine learning for reducing the cost such as a significant amount of time and domain expertise [80]–[82]. Bui et al. [83] proposed a flash prediction model using Particle Swarm Optimization (PSO) in an extreme learning machine that delivers the weights efficiently.

III. PROPOSED RISK INFERENCE APPROACH BASED ON GEOGRAPHICAL DANGER POINTS

The proposed risk inference approach is used for predicting risk index to be aware of dangers in SRTS based on geographical information. The inference approach is comprised of the RNN model, one-hot encoder, geohash converter, and data to predict the route risk index. The data for building the approach is collected from the Jeju province, South Korea. The data is provided by students through the survey that includes path, danger points, danger type, transportation, and gender. The dataset includes 1707 rows that are used for training and testing the RNN model separately to evaluate the performance.

Figure 1 presents the proposed data pipeline for the RNN-based inference approach. The data pipeline is comprised of input data, functional blocks, and RNN prediction models to predict the route risk index based on geographical information. For training the RNN model with the dataset, coordinates of routes and danger points are converted to be geohash through the geohash converter. The collected data is spatial data which is represented in two dimensional,

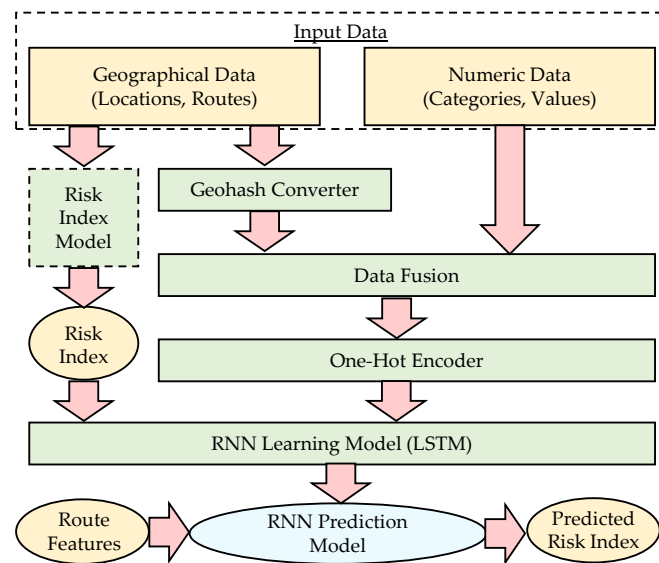


FIGURE 1. Proposed data pipeline for the inference approach of route risk index based on geographical information.

namely, longitude and latitude. This two-dimensional data is re-coordinate to single shorter string values using the geohash converter. Geohash algorithm provides a hierarchical grid-based model of the earth where locations are presented in Base32 strings [84]. To use more accurate, consistent data for our data processing, the converted geohash data, and other data features are integrated through data fusion. Data fusion helps to combine data from multiple sources or associated databases for improving accuracies and making simpler overall operational processes [85]. Fused data are inputted to the one-hot encoder. The one-hot encoding is a data pre-processing process in which categorical variables are converted to binary parameters which can provide better performance to ML algorithms. The one-hot encoded data is used for the inputs of the RNN model to train the LSTMs. The output of the trained model is derived by the risk index model that is proposed to calculate the risk index through route coordinates and danger points.

RNN is used for training sequence data to derive the next value in a real-valued sequence or outputting a class label for an input sequence. For providing the risk index in SRTS, the proposed approach can be categorized as the solution that processes multiple-input time steps to deliver one output time step. LSTM is a popular approach that achieves outstanding accuracy in training the sequences [34]. The provided data is converted to a string through the geohash converter and data fusion. However, the string data is comprised of characters that need to be converted to numerical data for training in the ML [86], [87]. Therefore, in the proposed approach, the one-hot encoding is applied to the integer representation.

Figure 2 depicts the risk index model dataflow. The risk index model includes two parameters, danger points, and route points. Each danger and route points' latitude and longitude parameters are respectively presented using $(D_{lat}(j), D_{lng}(j))$ and $(R_{lat}(i), R_{lng}(i))$. Latitude and longitude play

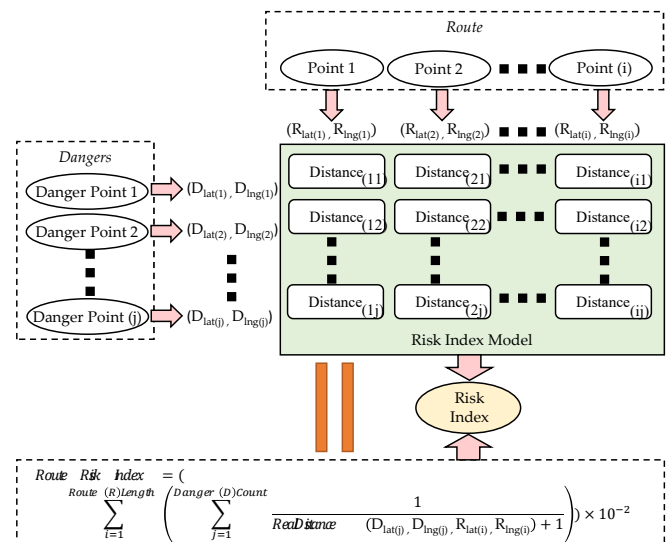


FIGURE 2. Risk index model dataflow.

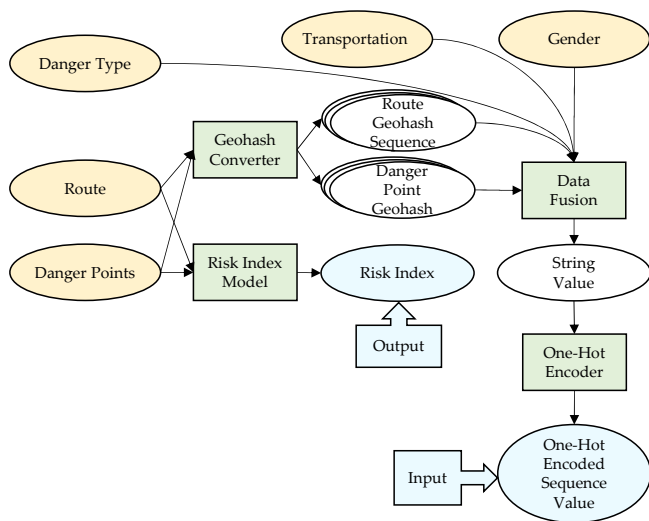


FIGURE 3. Figure 3. Data processing flow for inputs of the RNN learning model.

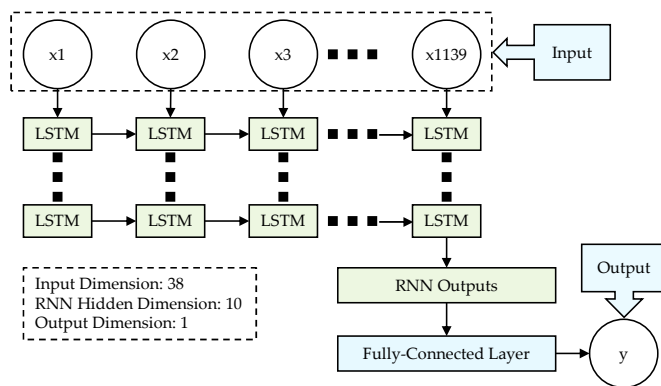


FIGURE 4. The architecture of the RNN learning model for one-hot encoded sequence value.

an essential role in several areas, but one of their most valuable usages is the measurement of distances between two locations. Especially, aerospace engineering, logistics, transportation and just to name a few. Distance calculations provide the shortest, the fastest and the most optimal routes between two locations [88]. Distance is calculated between all of the danger points and route points. The route risk index is calculated by the below-given equation. As can be seen equation, based on a real distance between danger points and route points. The route risk index is calculated by the below-given equation.

Figure 3 presents the data processing flow for RNN learning model input parameters. The first and very important step for conducting the route risk index analysis is the selection of input and output variables properly. Input variables are based on geographical information danger point coordinates of school routes danger points and risk levels which are collected by student survey. These data set coordinates are described in the two-dimensional format including latitude and longitude of locations. Output variables are risk in-

dex and risk level of routes. Route and danger points are converted to the geohash data using a geohash converter. Geohash data helps to encode latitude and longitude based geographic location into a short string of digits and letters. Geohash converter provides route geohash sequence and danger point data in geohash data format. The converted geohash data and other data of features including danger type, transportation, and gender of students are fused to the one data set. The fused dataset contains categorical variables such as gender, name of transportation, danger type, and geohash based danger points' location. These different data variables manner confuses the ML model, to avoid this the data should be encoded. One-hot encoded data represents each type of data in a format that the computer easily can understand through the binary format for the LSTM-based RNN.

Figure 4 presents one-hot encoded data based RNN learning model architecture. Then input and output parameters are presented as x and y for the learning module, respectively. RNN can provide one of the most powerful algorithms for processing sequential data. LSTM defines the unit of computation (memory cell) that replaces traditional artificial neurons in the hidden layer of the network. Networks can be able to associate with these memory cells effectively and high prediction performance can be achieved [89]. As we have mentioned above the converted geohash data and other data features are converted to the one-hot encoding data format. The input layer consists of 38 features that are each connected to all RNN hidden neurons (10) and the outputs are taken to the fully connected layer that outputs are the route danger points and route risk index.

IV. DATA PROCESSING AND PRESENTATION

For providing the data to the proposed inference model based on geographical information, we collect the data from students through a survey that includes path, danger points, danger type, transportation, and gender. Using data collection, feature selection, the geohash data converting, and data fusion, the data for a row of the dataset is translated to a string to be used for one-hot encoding to get a binary data. The original data is collected schools' route-ways and danger point coordinates from the Jeju province, South Korea. The dataset contains student survey-based danger points, students' gender, transportation type, and other features. Coordinates of location are given in latitude and longitude. For the original data, outliers, duplicates, and irrelevant records were identified and removed and other unimportant strings are cleaned for further processing.

The collected spatial data includes latitude and longitude of routes and danger points. These coordinates are given in the float data type. To improve the performance of the process based on latitude and longitude, the function of geohash data conversion is used for converting float data to geohash data and described with a short string of digits and letters. The data fusion is used for integrating the data from multiple sources to offer more compatible and accurate data.

No.	ID	Gender	TP	Route	DP1	DP2	DP3	DT1	DT2	DT3	DL
1	4551	M	1	MULT	POINT	POINT	POINT	31	31	31	3
2	4552	F	3	MULT	POINT	POINT		33	99		5
3	4553	M	3	MULT							5
...											
1705	8547	F	1	MULT	POINT	POINT	POINT	12	22		4
1706	8548	F	3	MULT							4
1707	8549	F	1	MULT							4

MULTILINESTRING((126.58983518605717 33.26810735108034)
 33.25218257470514,126.56474698353856 33.252108306330406,126.563782909076
 33.251762309062556,126.56317817061634 33.25252657186056,126.56389673807264
 33.25261024044725,126.5639143 33.2527179))

FIGURE 5. Original dataset of route information.

No.	ID	Gender	TP	Route	DP1	DP2	DP3	DT1	DT2	DT3	DL
1	4551	0	1	wvfj28m471d5 wvfj287buk	9	9	9	3			
2	4552	1	3	wvfj23vbfnoy wvfj23z1j6er	11	13		5			
3	4553	0	3	wvfj0z2f8kdk wvfj0z07etd				5			
...											
1705	8547	1	1	wvfj0jdkycuz wvfj0jd7n0m	2	6	9	4			
1706	8548	1	3	wvcvph6n7np wvcvpyfvk				4			
1707	8549	1	1	wvfj0j3vzn4n wvfj0j6h4857				4			

wvfj0j3vzn4n wvfj0j6h4857 wvfj0j3es672
 wvfj0j2ctcu8 wvfj0j2qpb1x wvfj0j3n2xc6
 wvfj0j3p0gys

FIGURE 6. Converted dataset of route information.

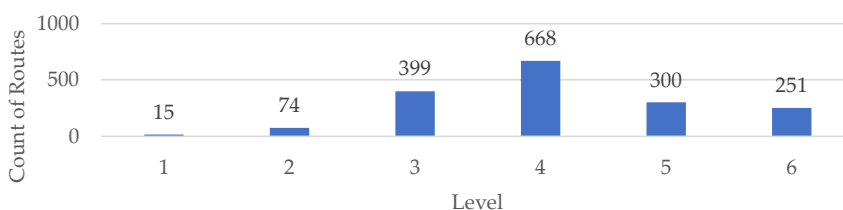


FIGURE 7. Count of routes for each danger level.

The converted geohash data and other data features fused and inputted to the one-hot encoder. One-hot encoding is a conversion process of categorical data to binary form which ML algorithms can do a better prediction performance. For improving our proposed prediction approach performance, all of the fused data is converted to the one-hot encoding based binary data. After data preparation processes, the one-hot encoded data is used to train RNN LSTMs. The outcome of the proposed model is route danger level and risk index.

Figure 5 presents the original dataset of route danger points

and the risk levels used for the proposed approach. The dataset includes several columns such as number, id, gender (Male or Female), transportation type (TP), danger points (DP), danger type (DT) and danger level (DL). As can be seen, the dataset includes 1707 routes for going to schools in Jeju. The most important part of this database is route danger point coordinates and risk levels. Danger points reveal the route coordinates in two-dimensional coordinate variables: latitude and longitude. Danger levels are based on the degrees of the risk from 1 to 6. Danger level 1 is accepted as the least

dangerous, whereas danger level 6 is accounted for the most crucial level for students.

Figure 6 presents the dataset after processing techniques that presents the gender of the students who attended the survey through encoded values 0 (male) and 1 (female). The geographical latitude and longitude coordinates of the route are converted to the geohash data. Geohash data is converted of the two-dimensional latitude and longitude float data to the one-dimensional geohash data.

Figure 7 illustrates the number of routes in terms of each danger level. As we mentioned above 1707 school route danger points and 6 danger levels considered to conduct this work. It can be seen that in the highest quantity of routes (668), the danger level is equal to 4. In contrast, the lowest number of routes (15 and 74) accounted for danger level 1 and danger level 2, respectively. The second most widely danger level was 3 and it was counted in 399 routes. The figures for 300 and 251 routes were responsible for danger level 5 and danger level 6, respectively, in the collected dataset. The largest proportion of danger level in the dataset was comprised of 39.13% (danger level 4) of overall danger levels, whereas the smallest shares of danger levels were 0.88% and 4.34% for danger level 1 and danger level 2, respectively. The second highest contribution for danger level was accounted for danger level 3 with 23.37%, according to the survey. There was little difference between the figures for danger level 5 and danger level 6, as the former contributes the third-highest percentage with 17.57%, whereas, the share of the latter was marginally lower (14.7%).

Figure 8 presents the risk index of each route in ascending order. The risk index is categorized according to the growing order from 0 to 3. Each route risk index calculation is related to the route danger coordinates. If the student lives close to school and the route, which he or she uses, has more danger coordinates, then the risk index can be higher. If the student lives far away from school and danger coordinates are less in the route, after that, the risk index of the route becomes lower. Until 727 sequence numbers risk index is nearly 0, this is followed by a continuous increase, and the highest risk index is equal to 1.636.

Figure 9 shows the map-based visualization of routes

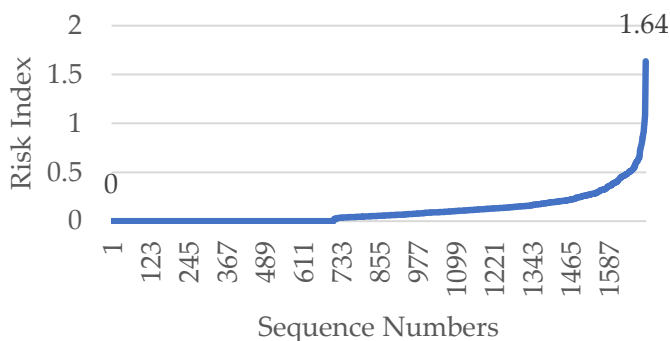


FIGURE 8. Risk index of each route.

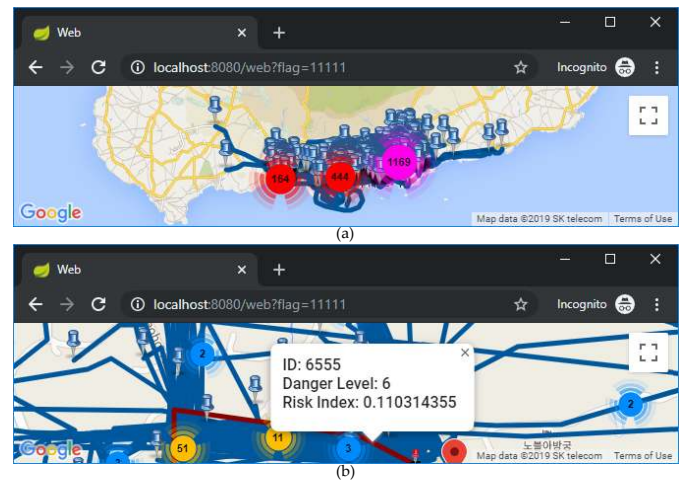


FIGURE 9. Map-based visualization of routes and danger points. (a) All routes and danger points. (b) Route information window for the clicked route.

and danger points. Map visualization helps to analyze and display the geographical information in the maps. This type of data description is clear and more understandable. Based on our prediction approach all routes and danger points are predicted using geographical information and presented on the map where 164, 444 and 1169 danger points are marked in separated groups. Once the route has been clicked from the window then the ID of the route and danger level and risk index can be visible.

V. EXPERIMENTAL RESULTS

Implementation of the proposed system consists of various steps and methods, such as data collection, data analysis, data conversion, data fusion, one-hot encoding, and RNN LSTM based on data training. We use geographical information-based datasets which include routes' danger points of ways to go to schools in Jeju province (Korea), as well as students survey-based route danger levels and danger points. The presented data is applied to the proposed inference approach that is comprised of risk index model, geohash converter, data fusion, one-hot encoder, and RNN prediction.

Figure 10 shows the experiment details of the proposed inference approach including functional blocks and data including inputs and outputs. The experiment is separated into two parts that are data preparation and data process. The original dataset is extracted from a survey that includes coordinates of danger points and routes which are translated to the geohash through the geohash converter. The geographical information is used for providing the inputs data and output data of the proposed learning model. For the inputs of the learning model, the geohash converter translates the geographical information of the original dataset and generates the geohash data. Also, other route features are converted to be numeric data. For the output of the learning model, the risk index model derives the risk index data that is calculated by the proposed equation. The equation derives the risk index based on geographical data. Therefore, there

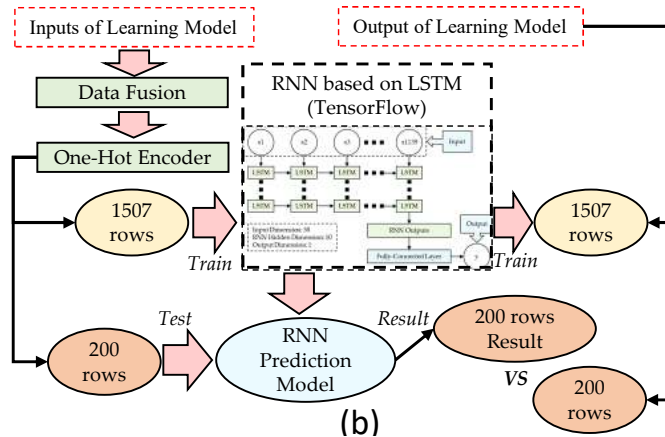
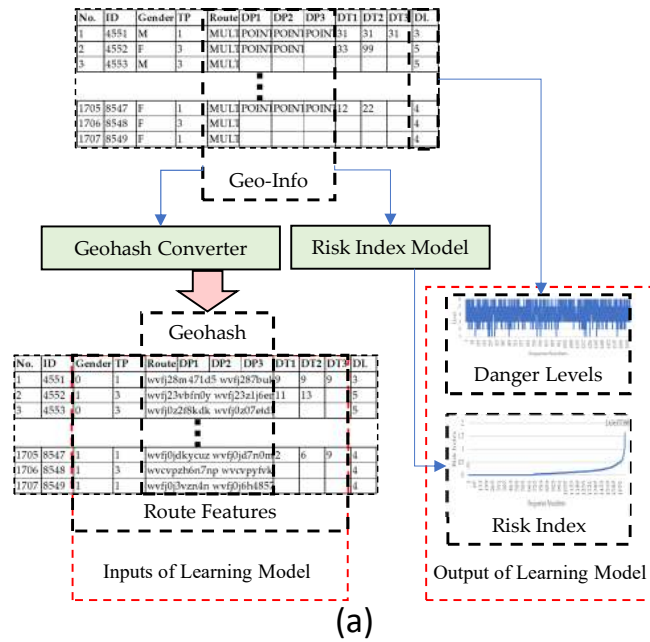


FIGURE 10. Experimental details. (a) Data preparation. (b) Data process.

is a reasonably high correlation between inputs data and risk index. Furthermore, we also bring danger levels to the leaning model for experimenting with the proposed learning model. However, the result of the inference model with the danger levels is not considered which can be assumed the danger levels are related to the geographical data.

The inputs of the learning model include data of gender, geohash and danger type 1,2, and 3. The translated geohash data is combined with a string for each row. Then, through the data fusion, all values of each row are fused into the on-hot encoder to prepare the training and testing dataset with a risk index for the RNN model that is implemented in TensorFlow 1.8. Also, instead of the risk index the danger level data also used in the RNN model. The total rows for training the learning model are 1507 that is comprised of strings. Each string presents the information of a record from the original dataset that includes geographical information and survey information. The RNN model is consists of LSTMs based on 10

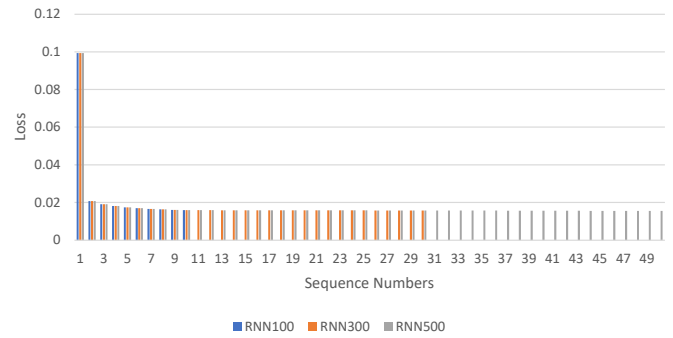


FIGURE 11. Loss over training epochs in RNN models for risk index.

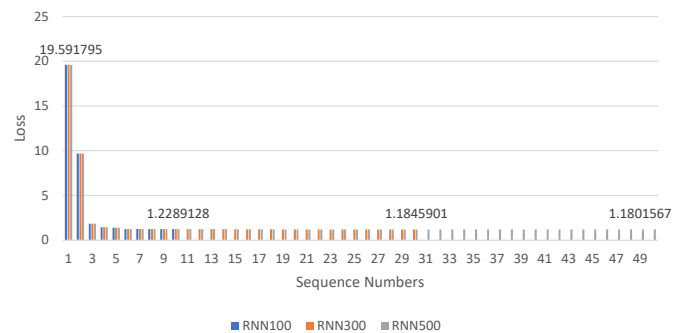


FIGURE 12. Loss over training epochs in RNN models for danger level.

hidden layers. The TensorFlow framework provides LSTM cells through the library to build the RNN architecture. At the last of the architecture, a fully-connected layer is added to deliver the result. Through training the RNN model with 100, 300, 500 epochs, the inference models are derived to predict the risk index. For testing the inference modes, 200 rows data are applied and evaluated using Mean Absolute Percentage Error (MAPE).

Figure 11 shows the loss over training epochs in RNN based risk index. Loss over training epochs measures the distance between the proposed risk index model’s output and the desired output during the training. It can be seen that in the initial steps of the training, the loss is about 0.1; after the second sequence numbers, the loss is decreasing gradually. There is no training error after the 11th and 30th sequence numbers in the case of RNN100 and RNN300, respectively. This means the proposed prediction is performing well.

Figure 12 shows the loss over training epochs in RNN based danger level prediction. Loss over training epochs measures the distance between the proposed danger level model’s output and desired output during the training. It can be seen that in the initial steps of the training, the loss is about 19.59 according to all RNN100, RNN300 and RNN500. After the 1.22 training loss, the RNN100 loss has changed to 0. RNN300 is changed to 0, in the sequence number 30, while the training loss is almost 1.84. The latest training loss is leveled off 1.180 based on RNN500 in sequence number 50.

VI. PERFORMANCE EVALUATION

For the risk index inference approach, the training data is comprised of 1507 rows. Each row includes information about gender, transportation, route, danger points, danger type, and risk index. The 1507 values of risk index are brought from the result of the risk index model to be used for the output, and others are the input which converted to one-hot encoded data. The RNN prediction model is generated by training the LSTM-based RNN model. We use 200 rows to test the RNN prediction model to the 200 rows result which is 200 values of risk index. From the result of the risk index model, we bring the corresponding 200 values to compare with the predicted values.

The geographical information-based data includes route and danger point coordinates in latitude and longitude values which are converted to a short string of digits and letters based geohash data using the geohash converter. Geohash converter converts two-dimensional coordinates (latitude and longitude) to one-dimensional value. One-dimensional geohash data is easy to implement complex scenarios. Converted geohash data and other data features are fused and encoded to the binary values (0 and 1) using One-Hot Encoder. The categories need to be converted into numbers using a one-hot encoder to achieve high performance for deep learning algorithms. One-hot encoded 1507 rows of data is used for training the RNN based LSTM model, 200 rows of data have been used to evaluation of the proposed prediction model. As well as, actual risk index and danger level (1507 rows) data applied to the RNN prediction model without using geohash converting, data fusion and one-hot encoding techniques.

Figure 13 compares the original risk index and predicted risk index results based on 100, 300 and 500 training epochs. As we have already mentioned above, as the number of training increases the error will also decrease. For risk index and danger level evaluation performance, we configure the RNN model with 100, 300, and 500 times training epochs to test the prediction performance. Epoch presents the process when the whole training data passes through the network, more precisely, an epoch is one iteration of the whole training data being passed through the network. For training our proposed model, we have split the training data into a batch size, where the batch size was defined as 100, 300 and 500. This means that the first 100 data have been taken from the training data (0-99) and trained on the network, after that it takes 100 samples (99-199) and trains the network. The same applies to 300 and 500 epochs as well. The epoch continues until all samples are propagated through the network hence then one epoch is passed through the network.

Figure 14 shows comparisons of the original danger level and RNN 100, 300 and 500 epochs based predicted danger level comparison results. Using the proposed inference approach, we applied a dataset that involves danger level data. With the dataset, the training model generates an RNN prediction model that is used for predicting the danger levels. However, according to the results, the data of danger level cannot be predicted well.

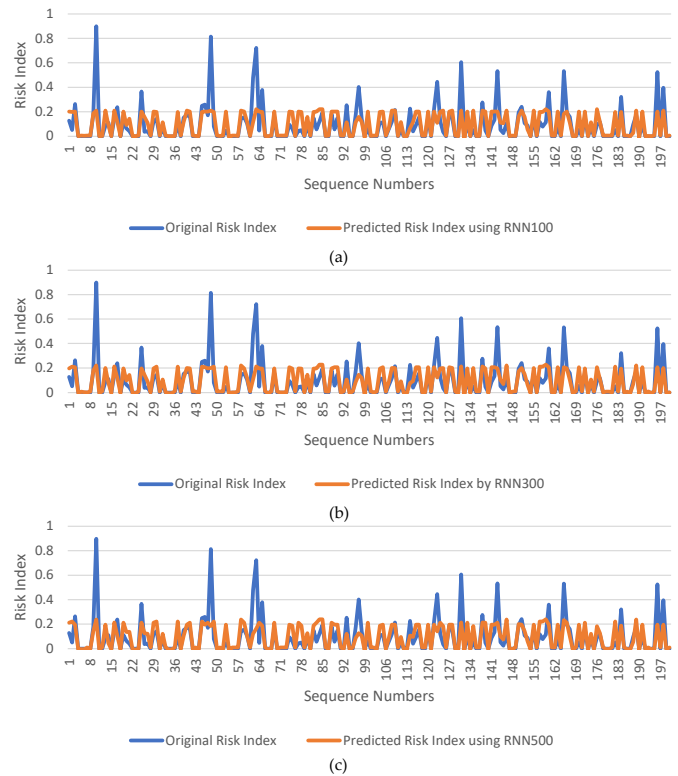


FIGURE 13. Comparison of original risk index and RNN-predicted risk index. (a) Training 100 epochs. (b) Training 300 epochs. (c) Training 500 epochs.

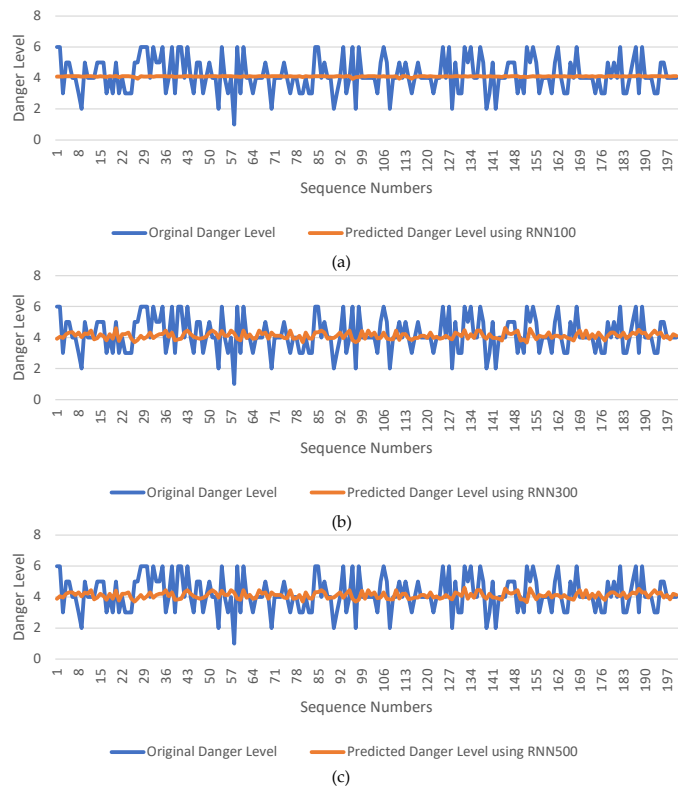


FIGURE 14. Comparison of original danger level and RNN-predicted danger level. (a) Training 100 epochs. (b) Training 300 epochs. (c) Training 500 epochs.

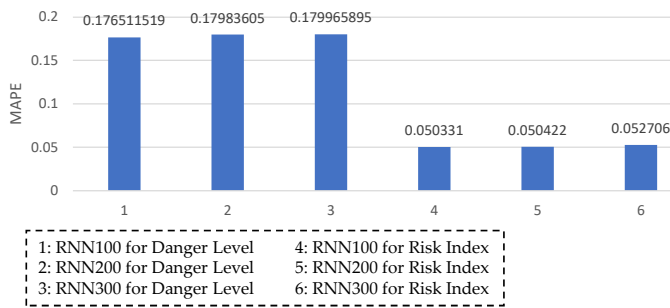


FIGURE 15. Comparison of evaluation results using MAPE

There are different ways to calculate the error or accuracy in the prediction models, such as Mean Squared Error, Root Mean Square Error, and MAPE. For comparing the experimenting results, we use MAPE to calculate the accuracy to evaluate the predicted results as shown in Figure 15. MAPE is a mathematical formula that gives us the ability to calculate the accuracy of our predicted risk index and predicted danger level. The calculation is done by taking the difference between actual values, predicted risk index, or predicted danger level dividing the difference by the actual value. In the next stage, it is multiplied by the number of data points and 100 to yield the percentage error. For predicting the risk index and danger level, the LSTM-based RNN model is trained with 100, 300, 500 epochs separately to get three different prediction models. For the results of the predicted risk index, MAPE values are calculated as 5.27%, 5.04%, and 5.03% using the three models. For the results of the predicted danger level, MAPE values are calculated as 17.99%, 17.98%, and 17.65% using the three models. The evaluation illustrates the proposed training model of RNN based on LSTM presents approximately the same performance by increasing training epochs.

For predicting the risk index, the result is much better than predicting the danger level. The reason can be assumed by the training data. The risk index in the training data is derived by the risk index model that calculates the risk index based on distances of data using the proposed equation. Therefore, the result of predicting risk index using the proposed inference approach is considerable because of the risk index correlating with other data from the training dataset.

VII. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we proposed an inference approach based on an RNN model with geographical information for the risk assessment of SRTS. The geographical information involves multiple coordinates that is difficult to be fused with other features for a one-hot encoded dataset to train an LSTM-based RNN model. For this purpose, we convert a set of coordinates to a string of geohash and combine with other information to make a string. Then we convert the string to a one-hot encoded data that improves the performance of the LSTM-based RNN model. For the implementation and experiment, the RNN models are configured to be 100,

300, 500 times training epochs to test the performance of the proposed inference approach using the geographical information. However, the MAPE-based evaluation illustrates the proposed inference approach presents approximately the same performance by increasing training epochs. The input data of the training model is derived by the risk index model that is proposed to calculate the risk index based on the student survey data. Therefore, the result of predicting risk index using the proposed inference approach is considerable because of the risk index correlating with other data from the training dataset.

As future directions, we will apply the proposed inference approach to multiple datasets such as path clustering, trajectory tracking, and other coordinates-based datasets. We will enhance the efficiency of the fused string data through removing and representing some unnecessary values such as space, and duplicated values of geohash to present significant results for learning to geographical information.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (NRF-2019R111A1A01062456), and this research was supported by Energy Cloud RD Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT (2019M3F2A1073387). Any correspondence related to this paper should be addressed to Dohyeun Kim.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests regarding the publication of this paper.

REFERENCES

- [1] D. A. Sleet and C. M. Branche, "Road safety is no accident," *Journal of safety research*, vol. 35, no. 2, pp. 173–174, 2004.
- [2] S.-P. Miaou, J. J. Song, and B. K. Mallick, "Roadway traffic crash mapping: a space-time modeling approach," *Journal of transportation and Statistics*, vol. 6, pp. 33–58, 2003.
- [3] Z. Jan, B. Verma, J. Affum, S. Atabak, and L. Moir, "A convolutional neural network based deep learning technique for identifying road attributes," in *2018 International Conference on Image and Vision Computing New Zealand (IVCNZ)*. IEEE, 2018, pp. 1–6.
- [4] H. Wang and Q. Liang, "Risk analysis and route optimization of dangerous goods transportation based on the empirical path set," *Journal of Advanced Transportation*, vol. 2020, 2020.
- [5] M. Holický, "Probabilistic risk optimization of road tunnels," *Structural Safety*, vol. 31, no. 3, pp. 260–266, 2009.
- [6] M. Stanković, Ž. Stević, D. K. Das, M. Subotić, and D. Pamučar, "A new fuzzy marcos method for road traffic risk analysis," *Mathematics*, vol. 8, no. 3, p. 457, 2020.
- [7] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives," *Transportation research part A: policy and practice*, vol. 44, no. 5, pp. 291–305, 2010.
- [8] R. B. Noland and M. A. Quddus, "A spatially disaggregate analysis of road casualties in england," *Accident Analysis & Prevention*, vol. 36, no. 6, pp. 973–984, 2004.
- [9] H. Lipson and M. Kurman, *Driverless: intelligent cars and the road ahead*. MIT Press, 2016.
- [10] V. Sumantran, C. Fine, and D. Gonzalez, "Innovations for intelligent cars and autonomy," 2017.

- [11] M. Gerla, E.-K. Lee, G. Pau, and U. Lee, "Internet of vehicles: From intelligent grid to autonomous cars and vehicular clouds," in 2014 IEEE world forum on internet of things (WF-IoT). IEEE, 2014, pp. 241–246.
- [12] N. C. McDonald, Y. Yang, S. M. Abbott, and A. N. Bullock, "Impact of the safe routes to school program on walking and biking: Eugene, oregon study," *Transport policy*, vol. 29, pp. 243–248, 2013.
- [13] N. C. McDonald and A. E. Aalborg, "Why parents drive children to school: implications for safe routes to school programs," *Journal of the American planning association*, vol. 75, no. 3, pp. 331–342, 2009.
- [14] M. Mehdizadeh, T. Nordfjaern, A. R. Mamdoohi, and A. S. Mohaymany, "The role of parental risk judgements, transport safety attitudes, transport priorities and accident experiences on pupils' walking to school," *Accident Analysis & Prevention*, vol. 102, pp. 60–71, 2017.
- [15] J. Parkin, T. Ryley, and T. Jones, "Barriers to cycling: an exploration of quantitative analyses," in *Cycling and society*. Routledge, 2016, pp. 83–98.
- [16] N. Soni and N. Soni, "Benefits of pedestrianization and warrants to pedestrianize an area," *Land Use Policy*, vol. 57, pp. 139–150, 2016.
- [17] F. Alonso, C. Esteban, F. Tortosa, and S. Useche, "Perception of road safety in children's environment," *Am. J. Educ. Res.*, vol. 5, pp. 273–278, 2017.
- [18] C. C. Reynolds, M. A. Harris, K. Teschke, P. A. Cripton, and M. Winters, "The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature," *Environmental health*, vol. 8, no. 1, p. 47, 2009.
- [19] M. M. Haque, H. C. Chin, and H. Huang, "Applying bayesian hierarchical models to examine motorcycle crashes at signalized intersections," *Accident Analysis & Prevention*, vol. 42, no. 1, pp. 203–212, 2010.
- [20] T. K. Anderson, "Kernel density estimation and k-means clustering to profile road accident hotspots," *Accident Analysis & Prevention*, vol. 41, no. 3, pp. 359–364, 2009.
- [21] N. Eluru, C. R. Bhat, and D. A. Hensher, "A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes," *Accident Analysis & Prevention*, vol. 40, no. 3, pp. 1033–1054, 2008.
- [22] R. Elvik, "The non-linearity of risk and the promotion of environmentally sustainable transport," *Accident Analysis & Prevention*, vol. 41, no. 4, pp. 849–855, 2009.
- [23] S. Chaudhuri, "Application of web-based geographical information system (gis) in e-business," in *Digital Marketing and Consumer Engagement: Concepts, Methodologies, Tools, and Applications*. IGI Global, 2018, pp. 649–665.
- [24] K.-T. Chang, "Geographic information system," *International Encyclopedia of Geography: People, the Earth, Environment and Technology*, pp. 1–10, 2016.
- [25] H. Senaratne, A. Mobasheri, A. L. Ali, C. Capineri, and M. Haklay, "A review of volunteered geographic information quality assessment methods," *International Journal of Geographical Information Science*, vol. 31, no. 1, pp. 139–167, 2017.
- [26] L. See, P. Mooney, G. Foody, L. Bastin, A. Comber, J. Estima, S. Fritz, N. Kerle, B. Jiang, M. Laakso et al., "Crowdsourcing, citizen science or volunteered geographic information? the current state of crowdsourced geographic information," *ISPRS International Journal of Geo-Information*, vol. 5, no. 5, p. 55, 2016.
- [27] H. Nguyen, L.-M. Kieu, T. Wen, and C. Cai, "Deep learning methods in transportation domain: a review," *IET Intelligent Transport Systems*, vol. 12, no. 9, pp. 998–1004, 2018.
- [28] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Briefings in bioinformatics*, vol. 19, no. 6, pp. 1236–1246, 2018.
- [29] M. Bakator and D. Radosav, "Deep learning and medical diagnosis: A review of literature," *Multimodal Technologies and Interaction*, vol. 2, no. 3, p. 47, 2018.
- [30] P. M. Addo, D. Guegan, and B. Hassani, "Credit risk analysis using machine and deep learning models," *Risks*, vol. 6, no. 2, p. 38, 2018.
- [31] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [32] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: A neural image caption generator," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3156–3164.
- [33] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3128–3137.
- [34] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.
- [35] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in *International conference on machine learning*, 2014, pp. 1764–1772.
- [36] L. Sun, J. Du, L.-R. Dai, and C.-H. Lee, "Multiple-target deep learning for lstm-rnn based speech enhancement," in *2017 Hands-free Speech Communications and Microphone Arrays (HSCMA)*. IEEE, 2017, pp. 136–140.
- [37] R. Messina and J. Louradour, "Segmentation-free handwritten chinese text recognition with lstm-rnn," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2015, pp. 171–175.
- [38] G. Jia, Y. Lu, W. Lu, Y. Shi, and J. Yang, "Verification method for chinese aviation radiotelephony readbacks based on lstm-rnn," *Electronics Letters*, vol. 53, no. 6, pp. 401–403, 2017.
- [39] Y. Lu, Y. Shi, G. Jia, and J. Yang, "A new method for semantic consistency verification of aviation radiotelephony communication based on lstm-rnn," in *2016 IEEE International Conference on Digital Signal Processing (DSP)*. IEEE, 2016, pp. 422–426.
- [40] S. Pascual and A. Bonafonte, "Multi-output rnn-lstm for multiple speaker speech synthesis and adaptation," in *2016 24th European Signal Processing Conference (EUSIPCO)*. IEEE, 2016, pp. 2325–2329.
- [41] J. Liu, H. Li, Y. Gao, H. Yu, and D. Jiang, "A geohash-based index for spatial data management in distributed memory," in *2014 22nd International Conference on Geoinformatics*. IEEE, 2014, pp. 1–4.
- [42] R. Moussalli, M. Srivatsa, and S. Asaad, "Fast and flexible conversion of geohash codes to and from latitude/longitude coordinates," in *2015 IEEE 23rd Annual International Symposium on Field-Programmable Custom Computing Machines*. IEEE, 2015, pp. 179–186.
- [43] G. Mboup, M. I. Camara, N. M. Balde, M. Sangare, and K. Fofana, "Towards a smart metropolitan regional development—spatial and economic design strategies: Conakry," in *Smart Metropolitan Regional Development*. Springer, 2019, pp. 239–310.
- [44] C. M. Winston, "A new route to increasing economic growth: Reducing highway congestion with autonomous vehicles," 2017.
- [45] S. M. Ciobanu and J. Benedek, "Spatial characteristics and public health consequences of road traffic injuries in romania," *Environ. Eng. Manag. J.*, vol. 14, pp. 2689–2702, 2015.
- [46] C. Wang, M. A. Quddus, and S. G. Ison, "Impact of traffic congestion on road accidents: a spatial analysis of the m25 motorway in england," *Accident Analysis & Prevention*, vol. 41, no. 4, pp. 798–808, 2009.
- [47] C. Chen, T. Li, J. Sun, and F. Chen, "Hotspot identification for shanghai expressways using the quantitative risk assessment method," *International journal of environmental research and public health*, vol. 14, no. 1, p. 20, 2017.
- [48] A. Moradi, H. Soori, A. Kavousi, F. Eshghabadi, E. Jamshidi, and S. Zeini, "Spatial analysis to identify high risk areas for traffic crashes resulting in death of pedestrians in tehran," *Medical journal of the Islamic Republic of Iran*, vol. 30, p. 450, 2016.
- [49] T. Steenberghen, T. Dufays, I. Thomas, and B. Flahaut, "Intra-urban location and clustering of road accidents using gis: a belgian example," *International Journal of Geographical Information Science*, vol. 18, no. 2, pp. 169–181, 2004.
- [50] V. Eksler and S. Lassarre, "Evolution of road risk disparities at small-scale level: example of belgium," *Journal of Safety Research*, vol. 39, no. 4, pp. 417–427, 2008.
- [51] M. S. Tehrani, B. Pradhan, and M. N. Jebur, "Spatial prediction of flood susceptible areas using rule based decision tree (dt) and a novel ensemble bivariate and multivariate statistical models in gis," *Journal of Hydrology*, vol. 504, pp. 69–79, 2013.
- [52] M. B. Kia, S. Pirasteh, B. Pradhan, A. R. Mahmud, W. N. A. Sulaiman, and A. Moradi, "An artificial neural network model for flood simulation using gis: Johor river basin, malaysia," *Environmental Earth Sciences*, vol. 67, no. 1, pp. 251–264, 2012.
- [53] M.-J. Lee, J.-e. Kang, and S. Jeon, "Application of frequency ratio model and validation for predictive flooded area susceptibility mapping using gis," in *2012 IEEE international geoscience and remote sensing symposium*. IEEE, 2012, pp. 895–898.
- [54] B. Pradhan, "Flood susceptible mapping and risk area delineation using logistic regression, gis and remote sensing," *Journal of Spatial Hydrology*, vol. 9, no. 2, 2010.

- [55] K. Chau, C. Wu, and Y. Li, "Comparison of several flood forecasting models in yangtze river," *Journal of Hydrologic Engineering*, vol. 10, no. 6, pp. 485–491, 2005.
- [56] A. K. Kar, A. Lohani, N. Goel, and G. Roy, "Rain gauge network design for flood forecasting using multi-criteria decision analysis and clustering techniques in lower mahanadi river basin, india," *Journal of Hydrology: Regional Studies*, vol. 4, pp. 313–332, 2015.
- [57] T. Tingsanchali and F. Karim, "Flood-hazard assessment and risk-based zoning of a tropical flood plain: case study of the yom river, thailand," *Hydrological Sciences Journal—Journal des Sciences Hydrologiques*, vol. 55, no. 2, pp. 145–161, 2010.
- [58] B. Merz, H. Kreibich, and U. Lall, "Multi-variate flood damage assessment: a tree-based data-mining approach," *Natural Hazards and Earth System Sciences (NHES)*, vol. 13, no. 1, pp. 53–64, 2013.
- [59] W. Deng, J. Zhou, Q. Zou, J. Xiao, Y. Zhang, and W. Hua, "Flood disaster evaluation model based on kernel dual optimization support vector machine," *Information Technology Journal*, vol. 12, no. 12, pp. 2412–2418, 2013.
- [60] M. S. Tehrani, B. Pradhan, and M. N. Jebur, "Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in gis," *Journal of hydrology*, vol. 512, pp. 332–343, 2014.
- [61] H. Mojaddadi, B. Pradhan, H. Nampak, N. Ahmad, and A. H. b. Ghazali, "Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and gis," *Geomatics, Natural Hazards and Risk*, vol. 8, no. 2, pp. 1080–1102, 2017.
- [62] A. Ebert, N. Kerle, and A. Stein, "Urban social vulnerability assessment with physical proxies and spatial metrics derived from air-and spaceborne imagery and gis data," *Natural hazards*, vol. 48, no. 2, pp. 275–294, 2009.
- [63] K. J. Mach, M. D. Mastrandrea, P. T. Freeman, and C. B. Field, "Unleashing expert judgment in assessment," *Global Environmental Change*, vol. 44, pp. 1–14, 2017.
- [64] Z. Vojinovic and M. B. Abbott, *Flood risk and social justice*. IWA Publishing, 2012.
- [65] M. Horritt, "A methodology for the validation of uncertain flood inundation models," *Journal of Hydrology*, vol. 326, no. 1–4, pp. 153–165, 2006.
- [66] G. Lecca, M. Petitdidier, L. Hluchy, M. Ivanovic, N. Kussul, N. Ray, and V. Thieron, "Grid computing technology for hydrological applications," *Journal of Hydrology*, vol. 403, no. 1–2, pp. 186–199, 2011.
- [67] N. A. Khan, "Real time predictive monitoring system for urban transport," Ph.D. dissertation, Kingston University, 2017.
- [68] Y. Lassoued, J. Monteil, Y. Gu, G. Russo, R. Shorten, and M. Mevissen, "A hidden markov model for route and destination prediction," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2017, pp. 1–6.
- [69] X. Zheng, D. Zhang, H. Gao, Z. Zhao, H. Huang, and J. Wang, "A novel framework for road traffic risk assessment with hmm-based prediction model," *Sensors*, vol. 18, no. 12, p. 4313, 2018.
- [70] S. A. R. Shah, T. Brijs, N. Ahmad, A. Pirdavani, Y. Shen, and M. A. Basheer, "Road safety risk evaluation using gis-based data envelopment analysis—artificial neural networks approach," *Applied Sciences*, vol. 7, no. 9, p. 886, 2017.
- [71] Y. Yang and M. Rosenbaum, "Artificial neural networks linked to gis for determining sedimentology in harbours," *Journal of petroleum Science and Engineering*, vol. 29, no. 3–4, pp. 213–220, 2001.
- [72] M. I. Sameen and B. Pradhan, "Severity prediction of traffic accidents with recurrent neural networks," *Applied Sciences*, vol. 7, no. 6, p. 476, 2017.
- [73] S. A. Ruk, S. G. Khan, S. Saddar, N. A. Mahoto, S. M. Z. Shah, and S. Brohi, "Evaluating route security and predicting the safest alternative using risk factors," *Indian Journal of Science and Technology*, vol. 11, p. 27, 2018.
- [74] Y. Guo, X. Jia, and D. Paull, "Effective sequential classifier training for svm-based multitemporal remote sensing image classification," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 3036–3048, 2018.
- [75] S. Jin, L. Wang, F. Guo, Q. Zou et al., "Aops-svm: a sequence-based classifier of antioxidant proteins using a support vector machine," *Frontiers in Bioengineering and Biotechnology*, vol. 7, p. 224, 2019.
- [76] M. Z. Islam, J. Liu, J. Li, L. Liu, and W. Kang, "A semantics aware random forest for text classification," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2019, pp. 1061–1070.
- [77] B. Xu, X. Guo, Y. Ye, and J. Cheng, "An improved random forest classifier for text categorization," *JCP*, vol. 7, no. 12, pp. 2913–2920, 2012.
- [78] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2224–2287, 2019.
- [79] H. Al-Sahaf, Y. Bi, Q. Chen, A. Lensen, Y. Mei, Y. Sun, B. Tran, B. Xue, and M. Zhang, "A survey on evolutionary machine learning," *Journal of the Royal Society of New Zealand*, vol. 49, no. 2, pp. 205–228, 2019.
- [80] L. Shao, L. Liu, and X. Li, "Feature learning for image classification via multiobjective genetic programming," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 7, pp. 1359–1371, 2013.
- [81] A. Agapitos, M. O'Neill, M. Nicolau, D. Fagan, A. Kattan, A. Brabazon, and K. Curran, "Deep evolution of image representations for handwritten digit recognition," in *2015 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2015, pp. 2452–2459.
- [82] L. Rodriguez-Coayahuitl, A. Morales-Reyes, and H. J. Escalante, "Structurally layered representation learning: Towards deep learning through genetic programming," in *European Conference on Genetic Programming*. Springer, 2018, pp. 271–288.
- [83] D. T. Bui, P.-T. T. Ngo, T. D. Pham, A. Jaafari, N. Q. Minh, P. V. Hoa, and P. Samui, "A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping," *Catena*, vol. 179, pp. 184–196, 2019.
- [84] A. Liu, Z.-X. Li, G.-F. Liu, K. Zheng, M. Zhang, Q. Li, and X. Zhang, "Privacy-preserving task assignment in spatial crowdsourcing," 2017.
- [85] F. Xiao, "Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy," *Information Fusion*, vol. 46, pp. 23–32, 2019.
- [86] P.-S. Huang, X. He, J. Gao, L. Deng, A. Acero, and L. Heck, "Learning deep structured semantic models for web search using clickthrough data," in *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, 2013, pp. 2333–2338.
- [87] X. Li, L. Peng, X. Yao, S. Cui, Y. Hu, C. You, and T. Chi, "Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation," *Environmental pollution*, vol. 231, pp. 997–1004, 2017.
- [88] B. W. Husted, D. Jamali, and W. Saffar, "Near and dear? the role of location in csr engagement," *Strategic Management Journal*, vol. 37, no. 10, pp. 2050–2070, 2016.
- [89] R. C. Staudemeyer and E. R. Morris, "Understanding lstm—a tutorial into long short-term memory recurrent neural networks," arXiv preprint arXiv:1909.09586, 2019.

...