Detection of Coordinate Based Accident-Prone Areas on Road Surface using Machine Learning Methods

Anitha Kumari Dara¹ and Dr. A. Govardhan²

¹ Research Scholar in CSE Department, JNTUH, Hyderabad, Telangana, INDIA

² Professor in CSE Department, JNTUH, Hyderabad, Telangana, INDIA

¹anithakumaridara@gmail.com, ²govardhan_cse@jntuh.ac.in

ABSTRACT

The growth in the road networks in India and other developing countries have influenced the growth in transport industry and other industries, which depends on the road network for operations. The industries such as postal services or mover services have influenced the similar growths in these industries as well. However, the dependency of these industries is high on the road surface conditions and any deviation on the road surface conditions can also influence the performance of the services provided by the mentioned services. Nonetheless, the conditions of the road surface are one of the prime factors for road safety and number of evidences are found, which are discussed in subsequent sections of this work, that the bad road surface conditions are increasing the road accidents. Several parallel research attempts are deployed in order to find out, the regions where the road surface conditions are not proper, and the traffic density is higher. Nevertheless, outcomes of these parallel works are highly criticised due to the lack of accuracy in detection of the road surface defects, detection of accurate location of the defects and detection of the traffic density data from various sources. Thus, this work proposes a novel framework for detection of the road defect and further mapping to the spatial data coordinates resulting into the detection of the accident-prone zones or accident affinities of the roads. This work deploys a self-adjusting parametric coefficient-based regression model for detection of the risk factors of the road defects and in the other hand, extracts the traffic density of the road regions and further maps the accident affinities. This work outcomes into 97.69% accurate detection of the road accident affinity and demonstrates less complexity compared with the other parallel research outcomes.

Keywords: *Correlation, Adaptive, Location Dependent, Accident Possibility Prediction, Regression.*

1. INTRODUCTION

The possibilities of the encountering an accident is rare and detecting the possibilities of encountering an accident is even rarer. However, detection of the possibilities of

encountering an accident can be very significant for various purposes such as making medical support available, generating the safe tracks for transportations and many more. Accident prone regions depends primarily on two aspects as road surface conditions and the traffic density. The first category must be detected, and the second category of the data must be fetched in order to detect the possibilities of encountering the accidents. The detected possible values for encountering the accident can lead to the avoidance of the accident for saving valuable human life.

In the recent past, a good number of parallel research attempts were made in order to detect the possibilities of the road accidents for making preventive measures and actions. Nevertheless, the outcomes of these parallel researches are highly criticized either for not being very accurate or for being highly time complex.

Majority of the parallel research outcomes have focused on the traffic data more rather than the road surface data as sequence of the events. This led to the lesser focus on the external influencing parameters such as road or weather conditions. Henceforth, motivated by the demand of the current research trends, this work formulates the following objectives for the research.

- Propose a machine learning method to predict the higher accident-prone regions
- Propose another machine learning method for avoidance of the accidents

The rest of the work is organized such that, in the Section – II, the parallel research outcomes are discussed, in Section – III, the mathematical model for the proposed solutions is discussed, in the Section – IV, the proposed algorithms are furnished, in the Section – V, the results obtained from the proposed framework is elaborated, in the Section – VI, the comparative analysis with the

parallel research outcomes are discussed, in the Section – VII the final research conclusion is presented.

2. PARALLEL RESEARCH OUTCOME

In order to formulate the problem in a better way, the understanding of the parallel research outcomes is highly crucial. Henceforth, in this section of the work, the parallel research outcomes are discussed.

As discussed in the previous section of this work, majority of the parallel research works have focused on the traffic pattern and mostly ignored the external influencing factors. In the same direction the work by Cai Y. et al. [1] have showcased the possibilities of traffic pattern modelling using the trajectories of the vehicles. This work is also criticised for being on the higher side of the framework complexity. Also, the work by Wang H. et al. [2] have aimed for similar outcomes with the reduced complexity of the model and aimed for detection of the next action on the event-based sequence of the modelling. In the other hand, additional to the previous two benefits, the work by Wang G. [3] have showcased that, the detection of the traffic sign is also an added information, that can help in predicting the possibilities of road accidents. This work is criticised highly for the higher complexity of the framework for making it adaptive for different road conditions. The bottlenecks of this work were addressed in the work by Khalid S. [4].

The other direction of this parallel research outcome has demonstrated that the information of the driving personals can also help in predicting the accident possibilities as showcased in the work by Yang A.Y. [5].

The mainstream research outcomes focus on the vehicle information such as distances between vehicles and the speed of the nearing vehicles. The work by Dong Z.S. [6] have showcased the optimal use of the sensors for these purposes. This work is widely accepted by many other parallel research attempts as the similar outcomes where produced by Mahmood Z. [7] for making the parking automation and avoidance of collisions during tight parking spaces. Nonetheless, the demand for using the machine learning methods for detection of the accident possibilities with the sense of severity analysis is the most effective way for achieving higher performance with lower complexity of the framework, which is also confirmed by many researchers as one of the notable outcome by Nguyen H. [8].

Henceforth, in the next section of the work, the problem solution, identified in this section of the work, is presented using the mathematical model.

3. MATHEMATICAL MODEL

After the detailed understanding of the gaps in the parallel research works, in this section of the work, the proposed solution is mathematically modelled.

Firstly, assuming that the complete dataset is identified as $DS[]$ and each and every tuple is denoted as D_x , then for a total number of n record, the relationship can be formulated as,

$$DS[] = \sum_{i=1}^n D_i \quad (\text{Eq.1})$$

Secondly, two parameters in the dataset are the coordinate parameters as (X,Y) . Thus, the from the initial set of parameters, these two parameters must be kept aside for the next phase of calculations. Henceforth, the new reduced dataset is assumed as $DS'[]$ for a total m number of elements.

Further, a new class variable must be added to the existing dataset, as C_x , for making the regression model to work. The relation can be formulated as,

$$C_x = \beta_0 + \beta_1.DS'[1] + \beta_2.DS'[2] + \dots + \beta_m.DS'[m] \quad (\text{Eq.2})$$

Where, the correlation coefficients are identified as β_x .

The correlation coefficients are calculated based on the inter-relation of the member of the tuples or the other attributes as,

$$\beta_x = \frac{\sum_{i=1}^m DS'[i]}{\sum_{i=1}^m DS'[i+1]} \quad (\text{Eq.3})$$

Where, “ i ” denotes subsequent attribute indexes from the datasets.

Further, the class variable C_x , is used to categorize the defects into four different classes, as,

$$C_x = \langle NoDefect, Patch, Crack, Pothole \rangle \quad (\text{Eq.4})$$

Finally, with the class variable and coordinate valuables are merged together to form the final dataset as $DS''[]$, as,

$$DS''[] = \langle C_x, X, Y \rangle \quad (\text{Eq.5})$$

And, with the highest density of the defects, the most accident-prone zones are identified.

Henceforth, based on the proposed solution model in this section of the work, the proposed algorithm is presented in the next section of this work.

4. PROPOSED ALGORITHM

After the mathematical presentation of the proposed solution, in this section of the work, the proposed algorithm is presented.

Algorithm: Correlation Based Adaptive Location Dependent Accident Possibility Prediction (**CBA-LD-APP**)

Step - 1. Accept the dataset as DS[]
Step - 2. For each DS[i]
 a. Separate the location coordinates as Corr[]
 b. Calculate the correlation coefficients, CC[],
 c. As, $CC[i] = DS[i]/DS[i+1]$
Step - 3. End
Step - 4. For each CC[j]
 a. For each DS[i]
 i. Calculate the class variable, $CD[i] = DS[i]*CC[j]$
Step - 5. End $CD_Fin[i] = \text{Max}(CD[i])$
Step - 6. For each DS[i]
 a. Group by Corr[j], For each Corr[j]
 i. If $(\text{Max}(CD_Fin))$ belongs to Corr[j]
 ii. Then, Identify Corr[j] Highly Accident Prone
 iii. Else, Identify Corr[j] less Accident Prone or Safe

b. End
Step - 7. End

In insights, connection or reliance is any measurable relationship, regardless of whether causal or not, between two irregular factors or bivariate information. In the broadest sense relationship is any measurable affiliation, however it regularly alludes to how much a couple of factors are directly related. Commonplace instances of ward wonders incorporate the relationship between the physical statures of guardians and their posterity, and the connection between the cost of a decent and the amount the customers are eager to buy, as it is portrayed in the alleged interest bend.

Connections are helpful because they can show a prescient relationship that can be abused by and by. For instance, an electrical utility may create less force on a gentle day dependent on the connection between power request and climate. Right now, is a causal relationship, since extraordinary climate makes individuals utilize greater power for warming or cooling. Be that as it may, when all is said in done, the nearness of a connection isn't adequate to induce the nearness of a causal relationship. Further, in the next section of the work, the obtained results from the proposed algorithms are discussed.

5. RESULTS AND DISCUSSIONS

After the detailed understanding of the parallel research outcomes with the research bottlenecks, mathematical formulations and furnishing of the proposed algorithm, in this section of the work, the obtained results are furnished and discussed.

Table 1: Firstly, the complete dataset is furnished here for initial visualization.

Dataset ID	Long Slope (Degree)	Perpendicular Slope (Degree)	Defect Number	Distance between the Defects (cm)	Depth (cm)	Average Width (cm)	Area (cm ²)	Latitude	Longitude	Severity of the Defect
1	-0.21	4.71	1	250.46	122.23	88.89	22263	17.70885659	77.40629339	None
2	0.33	-3	1	199.22	65.21	43.06	8579	17.52843918	77.48322928	None
3	-0.21	4.7	1	976.34	266.06	241.2	235494	17.66321757	77.41451252	None
4	-0.3	3.29	1	402.39	189.53	49.69	19994	17.68132596	77.44766173	Low
4	-0.15	6.78	2	246.64	182.18	36.94	9110	17.68132596	77.44766173	Low
4	0.23	-4.3	3	278.23	185.52	27.47	7644	17.68132596	77.44766173	Low
5	0.65	-1.54	1	166.98	99.18	55.89	9333	17.61788828	77.3871209	Low
6	0.72	-1.4	1	907.43	648.38	397.55	360747	17.77819666	77.84300895	Moderate
7	0.6	-1.67	1	602.82	692.67	167.89	101208	17.84221494	77.71907362	Moderate
7	0.78	-1.28	2	870.4	755	308.03	268105	17.84221494	77.71907362	Moderate
8	0.75	-1.33	1	1029.8	774.36	371.21	382272	17.76095889	77.56454795	High
9	0.17	-5.97	1	175.41	21.17	19.46	3413	17.69644945	77.36062153	High
10	0.45	-2.23	1	354.05	158.68	88.54	31347	17.70137112	77.83092853	None

The initial dataset is again analysed graphically here [Fig 1].

Fig. 1. Initial Dataset Analysis

Secondly, after the removal of the location parameters, the threshold calculation for remaining data attributes are performed and the results are furnished here [Table – 2].

Table 2: Threshold calculation Results

Parameter Names	Threshold Calculation Results
Long Slope	0.293076923
Perpendicular Slope	-0.249230769
Defect Number	1.307692308
Distance between the Defects	496.9361538
Depth	320.0130769
Average Width	145.8323077
Area	112269.9231

As the scale differs, maybe one is changing the amplifying intensity of a notional magnifying lens seeing the framework. In purported realizable speculations, the framework at one scale will by and large be believed to comprise of self-comparative duplicates of itself when seen at a littler scope, with various parameters depicting the segments of the framework. The obtained results are also visualized graphically here [Fig – 2].

Fig. 2. Threshold Calculation Results

Thirdly, the value for the coefficients are calculated and the obtained results are furnished here [Table – 3].

Table 3: Threshold calculation Results

Parameter Names	Coefficients	Coefficient Value
Long Slope	Coefficient. 0	0.000
Perpendicular Slope	Coefficient. 1	-1.176
Defect Number	Coefficient. 2	-0.191
Distance between the Defects	Coefficient. 3	0.003
Depth	Coefficient. 4	1.553
Average Width	Coefficient. 5	2.194
Area	Coefficient. 6	0.001

The polymorphic relationship coefficient estimates relationship between two arranged unmitigated factors. It's characterized as the gauge of the Pearson relationship coefficient one would acquire if the two factors were estimated on a persistent scale, rather than as requested class factors, and the two consistent factors followed a bi-variate typical circulation. At the point when the two factors are dichotomous rather than requested downright, the polymorphic relationship coefficient is known as the metaphoric connection coefficient. The obtained results are also visualized graphically here [Fig – 3].

Fig. 3. Coefficients Calculation Results

Fourthly, the Accident Possibility Detection results are furnished here [Table – 4].

Table 4: Accident Possibility Detection Results

Data set ID	Latitude	Longitude	Detected Defect Class	Accident Possibility Detection
1	17.70885659	77.40629339	0	Clear Road
2	17.5284391	77.48322928	0	Clear Road
3	17.66321757	77.41451252	1	Clear Road
4	17.68132596	77.44766173	2	Minor Patches, Low Accident Prune
4	17.68132596	77.44766173	2	Minor Patches, Low Accident Prune
4	17.68132596	77.44766173	2	Minor Patches, Low Accident Prune
5	17.61788828	77.3871209	2	Minor Patches, Low Accident Prune
6	17.7781966	77.84300895	4	Cracks on Road, Mid Accident Prune
7	17.84221494	77.71907362	5	Cracks on Road, Mid Accident Prune
7	17.84221494	77.71907362	6	Cracks on

				Road, Mid Accident Prune
8	17.76095889	77.56454795	9	Potholes on Road, High Accident Prune
9	17.69644945	77.36062153	9	Potholes on Road, High Accident Prune
10	17.70137112	77.83092853	2	Minor Patches, Low Accident Prune

The results are visualized graphically here [Fig – 4].

Fig. 4. Accident Possibility Analysis Zones

Finally, the accuracy analysis is carried out and the results are furnished here [Table – 5].

Table 5: Accident Possibility Detection Accuracy

Dataset ID	Accident Possibility Detection (Dataset Results)	Accident Possibility Detection (Proposed Algorithm Analysis Results)	Accuracy (%)	Overall Accuracy (%)
1	None	Clear Road	100	97.69
2	None	Clear Road	100	
3	None	Clear Road	100	
4	Low	Minor Patches, Low Accident Prune	100	
4	Low	Minor Patches, Low Accident Prune	100	

4	Low	Minor Patches, Low Accident Prune	100
5	Low	Minor Patches, Low Accident Prune	100
6	Moderate	Cracks on Road, Mid Accident Prune	100
7	Moderate	Cracks on Road, Mid Accident Prune	100
7	Moderate	Cracks on Road, Mid Accident Prune	100
8	High	Potholes on Road, High Accident Prune	100
9	High	Potholes on Road, High Accident Prune	100
10	None	Minor Patches, Low Accident Prune	70

Henceforth, it is natural to realize that, the proposed method demonstrates significantly high accuracy for the possibility detection of the accidents.

Furthermore, in the next section of the work, the proposed framework is compared with the other parallel research outcomes.

6. COMPARATIVE ANALYSIS

After the detailed analysis of the obtained results, in this section of the work, the parallel research outcomes are compared with the proposed framework, in order to obtain the confidence of being better compared with all other parallel research outcomes [Table – 6].

Table 6: Comparative Analysis

Method, Author & Year	Model Complexity	Features	Accuracy (%)
Trajectory-based, Y. Cai, 2015 [1]	$O(n * m)$	Traffic Surveillance	91
Dense Trajectory, H. Wang,	$O(n^m)$	Action recognition	95

2011 [2]			
Auto Parking, H. Mahmood, 2019 [7]	$O(n^3)$	Collision Avoidance	89
Proposed CBA-LD-APP Algorithm	$O(\log n)$	Road Surface Defect Detection, Severity Detection, Accident Prone Zone Marking	97.69

Henceforth, it is natural to realize that, the proposed method has outperformed the other parallel research works with a significant difference.

Furthermore, in the next section of the work, the research conclusion is presented.

7. CONCLUSION

With the growth of the surface transportation systems, the management and monitoring of the road network is gaining its importance. The major setback of further enhancement of any country relies primarily on the inter-country road transportation. It is often observed that the road surface conditions play major role in deciding the safety of the complete transportation system. However, majority of the parallel research outcomes have grossly ignored this factor and primarily focused on the traffic density and other traffic related parameters. Thus, this work identifies the bottlenecks of the parallel research outcomes and have incorporated the road conditions analysis for identification of the road accident possibilities. This work extracts the parameters from the geodetical road surface data and further processes the data and extracts the parameters like Long Slope (Degree), Perpendicular Slope (Degree), Number of Defects, Distance between the Defects (cm), Depth of the defect (cm), Average Width of the defect(cm) and the total Area of the defect(cm^3). Using these parameters, the work identifies the initial thresholds of the values and use the thresholds for determining the regression coefficients. Further, using the regression coefficients, the work again proposes a novel regression model to classify the defects into four major categories and finally maps the severity of the defect categories on the coordinates of the regions. The final outcome of this work is detecting the accident possibilities with 97.69% with lowest complexity of the model for making the road surface transportations faster and safer.

REFERENCES

- [1] Cai, Y., Wang, H., Chen, X., et al: 'Trajectory-based anomalous behaviour detection for intelligent traffic surveillance', IET Intell. Transp. Syst., 2015, 9, (8), pp. 810–816.
- [2] Wang, H., Klaser, A., Schmid, C., et al: 'Action recognition by dense trajectories'. Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 2011, pp. 3169–3176.
- [3] Wang, G., Ren, G., Wu, Z., et al: 'A robust, coarse-to-fine traffic sign detection method'. Proc. of Int. Joint Conf. on Neural Networks, Dallas, TX, USA, 2013, pp. 1–5.
- [4] Khalid, S., Muhammad, N., Sharif, M.: 'Automatic measurement of the traffic sign with digital segmentation and recognition', IET Intell. Transp. Syst., 2018, 13, (2), pp. 269–279.
- [5] Yang, A.Y., Zhou, Z., Balasubramanian, A.G., et al: 'Fast l1-minimization algorithms for robust face recognition', IEEE Trans. Image Process., 2013, 22, pp. 3234–3246.
- [6] Dong, Z.S., Zhu, W.X.: 'Homotopy methods based on l0-norm for compressed sensing', IEEE Trans. Neural Netw. Learn. Syst., 2017, 29, pp. 1–15.
- [7] Mahmood, Z., Haneef, O., Muhammad, N., et al: 'Towards a fully automated car parking system', IET Intell. Transp. Syst., 2019, 13, pp. 293–302.
- [8] Nguyen, H., Cai, C., Chen, F.: 'Automatic classification of traffic incident's severity using machine learning approaches', IET Intell. Transp. Syst., 2017, 11, (10), pp. 615–623.