A Mixed-Method Proposal for Traffic Hotspots Mapping in African Cities using Raw Satellite Imagery

Yilak A. Kebede

Lantern Institute of Data Science, Toronto, Canada

Abstract:-*Road traffic fatalities disproportionately affect low- and middle-income countries. This research provides a method that helps cities in developing countries to use their limited resource to control accident-prone locations with satellite data insights. The proposed method is a mixed approach from both transport and the emerging machine learning discipline. In the first step, accident spots labeled using the Weighted Severity Index (WSI) with 14 risk factors that potentially influence the occurrence of an accident. Then, the computer is trained to look for blackspots using the labeled geo-information data obtained from the WSI analysis. This cutting-edge method is called transfer learning with Convolutional Neural Networks (CNNs), which is the knowledge gained from previous training uses to identify a similar problem to a new location. The method is an inexpensive and reliable blackspot identifying solutions that extract data insights from freely available satellite imagery and open-source data.*

Keywords—road accident, hotspots; mapping; satellite imagery

I. INTRODUCTION (*HEADING 1*)

About 90% of global road deaths occur in low and middle-income countries. If these countries don't act quickly, they could lose between 7 and 22% of their potential GDP/capital growth over the next 24 years [1]. Rapid population growth and urbanization of African cities have brought a subsequent increase in motorization that has placed a heavy burden on road safety. Moreover, it is a leading cause of death for young people aged 15-29 [1].

The concept for the proposed method is founded on the assumption that crashes do not randomly occur across time and space; on the contrary, they are more likely to happen at specific occasions and locations e.g. urban intersections [2]. This accident-prone location, usually called blackspots or hotspots, covers where safety improvement measures are expected to have the greatest economic effectiveness [3]. Some studies show, focusing on Black spot road safety interventions could decrease the rate of accidents up to 28% [4]. However, this also is an uphill challenge for cities in developing countries as identifying/mapping blackspot requires accurate data, in which the collection and managing aspects of data are cumbersome and expensive. Therefore, this methodological proposal is novel as it provides an affordable (inexpensive and non-labor intensive) blackspot mapping method for developing cities using raw satellite data and open data sources. In particular, the method enables developing cities to prioritize and map accident-prone locations for deploying their limited human and financial resources.

II. PREVIOUS WORKS

In the literature, machine learning methodologies have been used for different urban studies application such as Arietta et al. [5] propose to predict location-aware crime rate and population from Google streetscapes based on visual pattern analysis, and learn prediction models with the visual patterns and the corresponding social attributes as input and output, respectively. Among recently conducted researches, Ameen.et.al [6] is a recent work that uses machine learning to predict city-scale road safety maps from raw satellite data. In this work, computational models are trained on satellite images mined from over 2.5 million official police reports collected by four different police departments in the US released as open data. The research conducted experiments on two different image classification architectures: (1) flat SVM-based architecture, and (2) deep ConvNet-based architecture. The Deep models outperform the flat models in predicting road safety from raw satellite imagery with an accuracy that reaches up to 79%. However, the research did not consider a more extreme case such as trying in a different geographical location such as targeting developing countries cities where the road facilities, city planning, and level of development extremely differed. Besides, the research used a general overview of injury frequencies rather than road crash Weighted Severity Index (WSI) to label the crashes. This reduces the amount of insight gained from the analysis that is useful to make an informed and learned decision by policymakers. Therefore, this proposed methodology extends further Ameen.et.al [6] work to predict and map road accident under different geographical truths, such as New York as a data source and African cities as a target of prediction.

III. METHODOLOGICAL PROPOSAL

The proposed method adopts the assumption taken to identify black spot from literature reviews: the site/area must have enough accidents to identify a pattern, the site/area must have enough accidents of the same type, and the accidents could be located on a map-based platform [7],[8].

This mixed-method approach is noble as it combines multidisciplinary methods from both transport and the emerging machine learning discipline. It is different from previous road safety mapping [6] by considering extreme cases such as training models with data from developed countries' sources to predict cities in developing countries as a target.

The method incorporates two separate methodological steps as shown in figure 1.

Fig. 1.   Methodological Steps
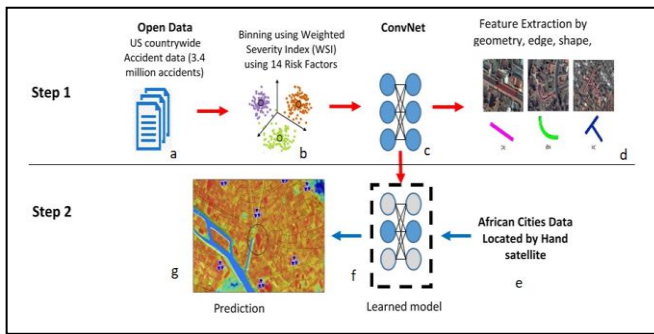
The methodological steps in their chronological order shown in *fig 1* as follows:

1. *Open-source data preparation*

Learning a computational model able to predict public safety from raw satellite imagery first requires collecting a set of training samples labeled with public safety. To obtain training data (labeled satellite images), the study proposes to mine large-scale collections of official police reports collected by police departments and released as open data.

Open data such as the US " A Countrywide Traffic Accident Dataset", which covers 49 states of the USA. The accident data are collected from February 2016 to June 2020 [12]. Currently, the dataset has about 3.5 million accident records. Each incident is described using attributes, such as time, date, geographic location, types of vehicle involved, and severity level.

Data collected from other available sources such as local police stations and insurance agencies are primary sources. The data should include at least four years of records and need to be digitalized using software such as Microsoft Excel or SPSS. The data should also have geometric coordinates, however, this might be difficult for some African cities, and they could fill the gap by using a handheld Garmin Global Positioning System (GPS). Also, photographs of the accident spots and interviews with all stakeholders, such as e.g. police, road users, traffic officials, are helpful to ascertain the black spots [8].

Image Labelling: - To obtain labeled satellite images from open datasets, the proposed method uses the Weighted Severity Index equation from step1. Sources, such as Google Static Maps API *https://developers.google.com/maps/documentation/static-maps*, could be used to crawl satellite images, in which individual images have a spatial resolution of 256 x 256 pixels each [6].

Resampling: - given that the obtained three classes are highly imbalanced and to avoid a biased model, the proposed method suggests resampling via down sample majority classes to balance out the three classes [8].

2. *Weight Severity Index analysis (Step1- b & Step 2- e in fig.1)*

WSI is a scientific method that is used for identifying the accident blackspots [9], based on their priority value criterion or distribution of Blackspots Severity (BSS). To understand the pattern of accidents, various spatial and non-spatial datasets should be collected, processed, and analyzed. WSI is based on the assumption that various factors influence the occurrence of road accidents. Based on their contribution, weights on a scale of 1-10 are assigned as shown in *Table 1*. Factors that tend to increase the probability of the accidents are assigned lower weights. *Equation 1* uses to calculate total weights.

TABLE 1: FACTORS USED IN THE WSI DETERMINATION, WITH POSSIBLE WEIGHTS (ADAPTED FROM OLUSINA.ET.AL 2017)

S/N	Factors	Possible Variation	Weight Assigned
F1	No. of lanes in each direction	1	2
		2	6
		4	10
F2	Width of Road	6-8m	3
		8-10m	5
		10-12m	7
		>12	10
F3	Type of road	access	2
		express	10
F4	Surface type	Bitumen	4
		Concrete	10
F5	Surface condition	Good	10
		fair	6
		poor	1
F6	Drainage facility condition	Good	10
		satisfactory	6
		poor	2
		Lack of availability	0
F7	Vehicle type	Heavy	10
		Vehicles	8
		Bus/ trucks	4
		Cars	2
		Motor cycles	1

S/N	factors	Possible Variation	Weight Assigned
F8	Number of vehicle per day	<10,000	10
		10,000-30,000	8
		30,000-50,000	4
F9	Shoulders	Paved	10
		Unpaved	6
		No	0
F10	Culvert/ Median #, Temporary barrier	yes	10
		no	4
F11	Traffic Control Mechanism	Good	10
		fair	6
		insufficient	4
		Not available	0
F12	Pedestrian Aid, movement across the road	Total pedestrian bridge movement	10
		Zebra line	6
		Avoid pedestrian aid	4
		Not available	0
F13	Freeness of human activities at shoulder	Totally free	
		Fairly cleared	
		Congested	
F14	Road design and shape	Straight	10
		Slope	8
		Sharp bend	6
		T-intersection	2
		Cross-intersection	1

The total weight for each road link should be calculated and normalized using the assigned maximum weight

$$Total\ Weight = (\textstyle\sum Wi)\ x\ 100/\ 140 \qquad (1)$$

Where $Wi = \sum$ individual weight of factors/ road segment

The road section with the highest value is less prone to accidents than the road section with a lower value [8]. The classification of the road section is done according to the values given in *Table 2*.

TABLE 2   PRIORITIZATION TABLE

Final Normalized weights	Accident Prone Level
75-100	Low
40-75	Medium
0-40	High

This prioritization method applies to label all accident locations based on accident-prone level. Once the data assessment is finished, geo-referencing of all accident blackspot locations should be recorded using hand-held GPS and transferred to the map using different data visualization tools and the data should be ready for the step ConvNet.

3. *ConvNet Model Learning*

The ConvNets is widespread and has been used in recent years for handling a variety and complex problems such as image recognition and classification by using a sequence of feed-forward layers designed to process data that come in

multiple arrays, such as RGB color images [6],[10]. ConvNets takes a raw RGB image as an input and produces a class prediction as an output.

A typical structure of CNN is a series of layers including a convolutional layer, a pooling layer, and full connection layers [10]. As *fig. 2* shows ConvNet has two parts: feature learning (Conv, Relu, and Pool) and Classification (FC and softmax). As *fig.2* shows, a typical convolutional layer convolves a three-dimensional input tensor with a tensor of weights (filter maps). The weighted sum of the convolution is then passed through a nonlinearity function such as a Rectified Linear Unit (ReLU). The result is then passed through pooling operators to reduce the dimensionality of the representation and make it invariant to small perturbations. On the other hand, a fully-connected layer reduces the multidimensional input into a one-dimensional vector that is fed to a final classifier.

Fig. 2.   A framework of a Convolutional Neural Network (Illustration by Mathworks)
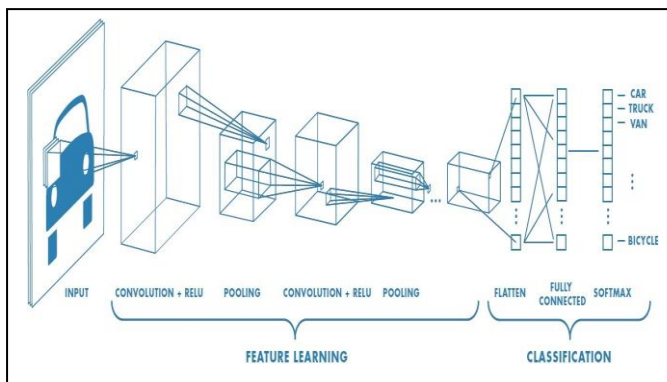
In all these layers there are only a few layers within CNN architecture that can be suitable for feature extraction of the input image. The features that have been extracted from the deeper layer can be used as a training feature because it gives advance features contrariwise the beginning layer of the CNN capture only the primary image features like edge and BLOBs (Binary Large OBject). The first layer of the CNN has learned for detecting the edge and blob features, and these original features are processed by a deeper layer in this case the first features are combined with more in-depth high-level features in the full connections layer, that can be used in recognition or classification tasks, so the fully connected layer is chosen to feature' layer [6],[10].

This paper suggests using and compare four CNN pre-trained models (AlexNet, VGG19, GoogLeNet, and Resnet50) for feature extraction, each of them trained on ImageNet dataset. These approaches enable us to combine the earlier features with more in-depth features in a fully connected layer and compare all the results of the models using a sample dataset from US countrywide accident data [11].

a.   Blackspot mapping proposal

The main assumption considered in this proposed method that mapping accident-prone location is a supervised image classification problem in which a city-scale satellite map is treated as a set of high-resolution satellite images. This method assumes that satellite imagery has a rich medium of visual features that can be used as a proxy indicator to extract hidden insight into the nature of accidents.

The overall framework is shown in *fig.3* for the automatic mapping of road crashes from satellite imagery. Two cities, City- A and City-B, serves as a source and target cities, and the goal is to generate for the target city a city-scale map indicating road safety in three different levels (low, medium, and high) using the Weight Severity Index, and predicted from its raw satellite imagery using minimal ground truth from city B.

Fig. 3.   ConvNet Approach adopted from [6]
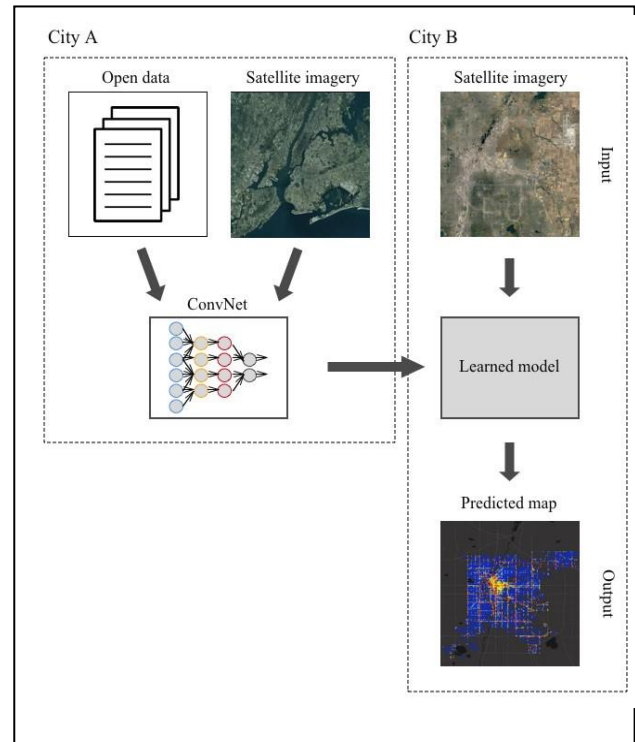
a.   Training and Testing

In this step, the proposed method has two parts: the training phase and the testing phase [10]. The datasets are divided into two sets initially the first is used as a training image and the second one used for testing the models.

Training: the proposed models initialized from the labeled image of US countrywide accident datasets. Training could be done using Caffe library framework[1] run on a single Nvidia GeForce TITAN X GPU. Caffe is helpful in vectorizing inputs data through special vectorization called blobs. A Blob is a type of array that speed up data analysis ability and provide synchronization ability between CPU and GPU.

Testing: The second phase of the satellite image classification model is a testing phase. In this part, the 30% remaining of each dataset should be tested to check and measure the accuracy of the classifier method [10].

Evaluation: To evaluate the learned models, this paper proposed the average prediction accuracy cross-validated on three random 5%/95% data splits. Reported results could be obtained after ~ 60,000 training iterations or more.

4. Tranfer learning (Step 2-f in fig.1)

The proposed method train models by adopting transfer learning in which pre-learned knowledge (from *step 1-c*) is

---

[1] https://caffe.berkeleyvision.org/

transferred from a source to a target problem, selectively some layers should be frozen. The transferred knowledge is a set of low-level visual features such as edges and corners. In the deep learning community, this way of training is known as fine-tuning and it has been proven highly successful in augmenting learning when training data is limited [6]. Therefore, to fine-tune a pre-trained model, first, this method replaces the classification layer with a three-class output layer representing the three safety levels [6]. Weights of the newly added layer initialized randomly, and the entire network should be trained jointly using small learning rates.

Fine-tuning: A large amount of data is needed to build a functional convolutional neural network model. In practice, it is common to reuse a pre-trained network. However, most pre-trained networks work for a different set of labels and were not trained with satellite images, so for this use case, it is necessary to retrain some of the top layers to improve prediction. The proposed methodology suggests trying different fine-tuning strategies shown in *fig.4* to fine-tune a pre-trained network with satellite imagery. The procedure is as follows: 1. Instantiate the convolutional base of a pre-trained model such as ResNet-50. 2. Add a fully-connected model on top, with a standard SGD optimizer and validating with the binary cross-entropy loss function. 3. Freeze the layers of the model up to the top 70 layers. 4. Retrain the model. This paper recommends the use of Keras library for data augmentation, training, and prediction [1]

To enhance the performance of the model, the research uses the small Addis Ababa's or other African cities dataset identified in Step1 to train the pre-trained model. Further, the proposed method should validate the model by using other Ethiopian and African cities' satellite images.

Fig. 4. Fine-tuning strategies [2]
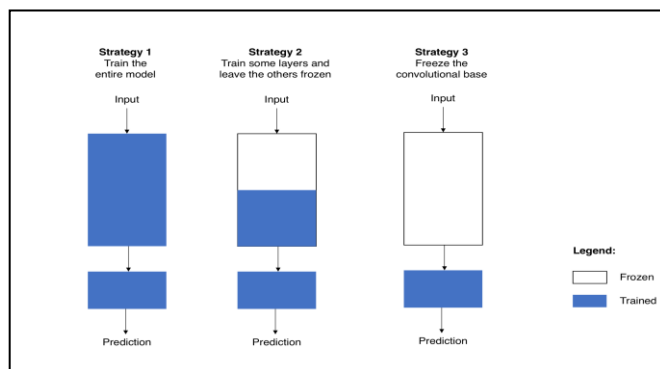
IV. CASE STUDIES

Case study 1- Road safety prediction from satellite imagery [6]

This project is among the recent attempts to investigate the visual features captured in satellite imagery that can be effectively used as a proxy indicator for road safety and to evaluate the ConvNet learning performance to predict road safety location from raw satellite imagery.

The study fine-tuned the ConvNet on images of the New York dataset. *Table 3.* shows the average prediction accuracy of nine models obtained considering three pre-training

---

² https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751

scenarios and using satellite images captured at three zoom levels

TABLE 3 AVERAGE PREDICTION ACCURACY OBTAINED USING NINE MODELS PRE-TRAINED ON THREE DIFFERENT LARGE-SCALE DATASETS AND FINE-TUNED ON SATELLITE IMAGES CAPTURED AT THREE DIFFERENT ZOOM LEVELS [6]

	x18	x19	x20
ImageNet	0.740	0.766	0.739
Places205	0.755	0.775	0.745
ImageNet + Places205	0.778	0.782	0.771

Lessons obtained:

1. For all zoom levels, models pre-trained on both ImageNet and Places205 achieve the best, followed by models pre-trained on Places205, and finally, models pre-trained on ImageNet. This is expected since satellite images have bird's eye/aerial viewpoint which makes them closer in composition to scene images of Places 205 rather than the object-centric images of ImageNet.

2. For all pre-training scenarios, fine-tuning using satellite images captured at zoom level x19 results in the best performance.

Results obtained in this experiment confirm the initial assumption that visual features captured in satellite imagery can be effectively used as a proxy indicator of road safety. Moreover, ConvNets are able to learn robust models that can predict road safety from raw satellite images.

The research also conducts a similar experiment on the Denver dataset, to investigate the reusability of the above learned deep model across different cities. Compared to the official accident map, the predicted map has an accuracy of 73.1%.

*Case study 2- Accelerating geospatial deep learning pipeline with fine-tuning [12]*

The study starts with the following questions: Can models be fine-tuned on the imagery of cities that they have never seen before? Is this process more efficient than training models from scratch or ImageNet weights, as was done previously?

The project used "the Solaris Versin of XD_XD's 5yh place model from SpaceNet 4, originally trained to identify buildings in imaginary of Atlanta, and fine-tuned it to perform the same task on the imagery of Khartoum, Sudan.

As seen in figure 3, the model trained to find buildings in the imagery of Atlanta really couldn't perform the same task in the imagery of Khartoum. This is common in deep learning models for computer vision — they can't perform well on imagery that's very different from anything they've ever seen before, a task termed "generalization". The study experimented further by fine-tuning the model — re-train it at a lower learning rate on new data for just a few epochs — and improve building footprint extraction quality in the imagery of Khartoum. The main objective of why they conducted the study was to reduce the learning rate 10-fold below what was used originally for training, and trained for three epochs on the SpaceNet 2 Khartoum dataset.

The result was successful, the fine-tuned model performed well compared to the original model.

Fig. 5. Prediction improvement after fine-tuning adopted from [12]
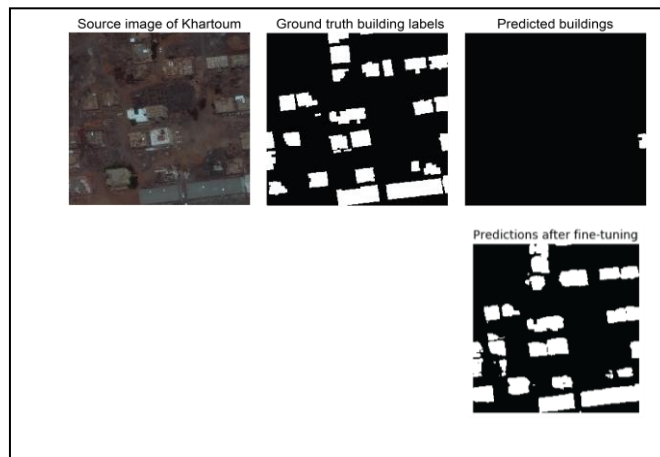
In *fig. 5*, a comparison of the original image of Khartoum (top left), the hand-labeled ground truth (middle), the predicted buildings before fine-tuning (top right), and the predictions after fine-tuning (bottom right). Though still imperfect, the buildings identified after fine-tuning are markedly better than before fine-tuning. This image was held out from the training set during fine-tuning.

V. DISCUSSION

This paper formalizes the problem of public safety mapping as a supervised image classification problem, in which a city-scale satellite map is treated as a set of satellite images each of each is assigned a safety label predicted using a model learned from training samples. To obtain this training data, the study suggests leveraging official police reports collected by police departments and released as open data. The idea is to mine large-scale datasets of official police reports for high-resolution satellite images labeled with safety scores calculated based on the number and severity/category of incidents. Then, validate and test the robustness of the learned models for road safety prediction tasks over four different US cities such as New York, Chicago, San Francisco, and Denver. At this stage, it is possible to investigate the reusability of the learned computational models across different cities.

From case study 1- the best performing models can predict road safety from raw satellite imagery with an accuracy that reaches up to 79%. Models learned from data collected in one city can be effectively (to a certain degree) reused across different cities. These results prove the assumption that visual information contained in satellite imagery has the potential to be used as an effective proxy indicator of public safety. Moreover, results obtained in case study 1 & 2 confirm that deep models learned from road safety data collected in a large city can be reused to predict road safety in smaller cities with fewer resources.

The main contributions made in this paper can be summarized as follows: (1) proposing a framework for automatic city-scale public safety prediction from satellite imagery, (2) proposing an automatic approach for obtaining labeled satellite imagery via mining large-scale collections of official police reports released as open data, and (3) introducing five labeled satellite imagery datasets representing different cities official police reports.

VI. CONCLUSION

Many African cities have had resource limitations to collect accident data, which is both expensive and labor-intensive. Therefore, this paper provides a methodological proposal to use open data and the visual information contained in raw satellite imagery as a proxy indicator for road safety. To fill the gaps, transfer learning is a good option for researchers and policymakers who have data and computational resource limitation but needs caution in avoiding misleading results. This could be done gradually and step-by-step experiments of fine-tuning parameters to get a clue on the trend of improvement. Utilizing pre-trained models for feature extraction coupled with selected classification methods, such as SoftMax could give great results. For feature work, the author is considering trying different ground truth data from different cities on pre-trained models and find a method to improve the classification accuracy when the source and target cities' geographical, architectural/planning contexts are different. Such road Safety analysis can be used to find out factors influencing crashes and hence to give remedial measures.

ACKNOWLEDGMENT

REFERENCES

[1] WHO, Global status report on road safety 2018. Geneva: World Health Organization; 2018. Licence: CC BYNC- SA 3.0 IGO.

[2] Marco Dozza , What is the relation between crashes from crash databases and near-crashes from naturalistic data?, Journal of Transportation Safety & Security, 12:1, pp. 37-51, 2020.

[3] Nguyen, Huy & Taneerananon, Pichai & Luathep, Paramet. (2016). Approach to Identifying Black Spots Based on Potential Saving in Accident Costs. Engineering Journal. 20. 109-122. 10.4186/ej.2016.20.2.109.

[4] Dhule, Jitesh & Gawade, Rutuja & Pawar, Amol & Borge, Swapnil & Ansari, Nazim & Sayyed, Shadaab. (2017). Accident Analysis & Blackspot Identification from Chandani Chowk to KJEI Campus, Pune- A Case Study. International Journal of Innovative Research in Science, Engineering and Technology. 6. 10.15680/IJIRSET.2017.0605276.

[5] Arietta, Sean & Efros, Alexei & Ramamoorthi, Ravi & Agrawala, Maneesh. (2014). City Forensics: Using Visual Elements to Predict Non-Visual City Attributes. IEEE Transactions on Visualization and Computer Graphics. 20. 2624-2633. 10.1109/TVCG.2014.2346446.

[6] Alameen Najjar, Shun'ichi Kaneko, Yoshikazu Miyanaga. (2017). Combining satellite imagery and open data to map road safety, AAAI Conference on Artificial Intelligence, pp. 4524-4530

[7] Tabatabaee, S.A. and Johari B. (2012). Providing a Model in order to Transportation Safety in Accident Black Spots in Suburban Roads from Economical Optimizing Point of View. Vol 4. IJET.

[8] OLUSINA, JO, AJANAKU, WA, "Spatial Analysis of Accident Spots Using Weighted Severity Index (WSI) and Density-Based Clustering Algorithm", J. Appl. Sci. Environ. Manage. April. 2017 Vol. 21 (2) 397-403

[9] B. Sudharshan Reddy, L. Vishnu Vardhan Reddy, G. Sreenivasa Reddy (2017), Statistical Analysis And Treatment Of Accident Black Spots: A

Case Study Of Nandyal Mandal, IOP Conference Series, Volume 225, 3-4 july 2017

[10] Kadhim, Mohammed & Abed, Mohammed. (2020). Convolutional Neural Network for Satellite Image Classification. 10.1007/978-3-030-14132-5_13.

[11] Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. "A Countrywide Traffic Accident Dataset.", 2019.

[12] Nick Weir, "Accelerating your geospatial deep learning pipeline with fine-tuning:- using CosmicQ's Solaris library to fine-tune models pre-trained on SpaceNet overhead imagery", Sept 11, 2019, https://medium.com/the-downlinq/accelerating-your-geospatial-deep-learning-pipeline-with-fine-tuning-248aa3cd4e89