

**SHORT-TERM SPATIOTEMPORAL CRASH PREDICTION  
AND MODELING – APPLICATION OF DEEP LEARNING  
AND HYBRID MODELS**

BY

**MOHAMMAD TAMIM KASHIFI**

A Thesis Presented to the  
DEANSHIP OF GRADUATE STUDIES

**KING FAHD UNIVERSITY OF PETROLEUM & MINERALS**

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the  
Requirements for the Degree of

**MASTER OF SCIENCE**

In

**CIVIL AND ENVIRONMENTAL ENGINEERING**

**APRIL 2021**

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

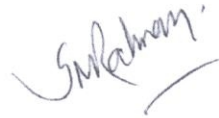
DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by Mohammad Tamim Kashifi under the direction of his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN CIVIL ENGINEERING.**



Dr. Hassan Mousaid Al-Ahmadi  
(Advisor)



Dr. Syed Masiur Rahman  
(Member)



Dr. Salah Othman Al-Dulaijan  
Department Chairman



Prof. Suliman Saleh Al-Homidan  
Dean of Graduate Studies



Dr. Ibrahim Yousef Saleh Al-Sghan  
(Member)

\_\_\_\_\_  
Date



© Mohammad Tamim Kashifi

2021

## **DEDICATION**

*I would like to dedicate this research to my parents, who always supported me unconditionally.*

## ACKNOWLEDGMENTS

First and foremost, I would like to thank Allah for giving me the strength and capability to complete my master thesis.

I would like to extend my sincere gratitude for the unconditional support and encouragement that I have received from my family. They have always been supporting me with their great counsel and encouragement. I would like to express my sincere gratitude to my wonderful mother, Mrs. Sabria Kashifi, and my great father, Mr. Eqbal shah Kashifi, for their continuous and unconditional support throughout my life. They have always stayed with me, and my studies would not have completed without their continuous support. I am incredibly grateful for having a family like you.

I am also thankful to my advisor Dr. Hassan Mousaid Al-Ahmadi. You have been supporting me throughout the completion of my thesis. I always appreciate your guidance and support that facilitated this research. Additionally, I must greatly appreciate the excellent guidance, support, and patience of my committee Dr. Syed Masiur Rahman. Your insightful feedback and immense knowledge helped me to cross the boundaries of my thoughts and elevate this research to its current level. I would never forget the lessons that I learned from you and the inspiration that you give me. I also thank Dr. Ibrahim Yousef Saleh Al-Sghan, for being on my committee and for his support.

In the end, I would like to thank all my friends and colleagues who inspired me and assisted me throughout the completion of this study.

# TABLE OF CONTENTS

ACKNOWLEDGMENTS.....	V
TABLE OF CONTENTS.....	VI
LIST OF TABLES.....	IX
LIST OF FIGURES.....	X
LIST OF ABBREVIATIONS.....	XII
ABSTRACT.....	XIV
ملخص الرسالة.....	XVI
CHAPTER 1 INTRODUCTION.....	1
1.1 Study Area.....	3
1.2 Objectives.....	4
1.3 Expected outcomes and their utility .....	5
1.4 Thesis Organization .....	6
CHAPTER 2 LITERATURE REVIEW.....	8
2.1 Crash Modelling and Prediction Methods.....	8

<b>2.2</b>	<b>Crash Spatial Aggregation for Modelling and Analysis .....</b>	<b>10</b>
<b>2.3</b>	<b>Crash Spatial Aggregation for Modelling and Analysis .....</b>	<b>11</b>
<b>2.4</b>	<b>Real-time crash prediction on Freeways and Expressways .....</b>	<b>12</b>
<b>2.5</b>	<b>Real-time crash prediction on intersections and weaving sections .....</b>	<b>13</b>
<b>2.6</b>	<b>Machine Learning Method to Predict Real-time Spatiotemporal crash grid-based.....</b>	<b>15</b>
<b>2.7</b>	<b>Machine Learning Methods to Predict Real-time Spatiotemporal crash grid-based .....</b>	<b>20</b>
	 <b>CHAPTER 3 METHODOLOGY.....</b>	 <b>22</b>
<b>3.1</b>	<b>Problem Setup .....</b>	<b>22</b>
3.1.1	Inputs .....	23
3.1.2	Desired .....	26
3.1.3	Output.....	26
3.1.4	Purpose .....	27
<b>3.2</b>	<b>Analysis Framework.....</b>	<b>27</b>
3.2.1	Data .....	29
3.2.2	Discretization to grids.....	43
3.2.3	Data Split .....	46
3.2.4	Data Preprocessing.....	46
3.2.5	Performance Measures .....	47
3.2.6	Model Selection and Training.....	48
3.2.7	Validation and Hyperparameter Tuning.....	53
3.2.8	Testing the models .....	53
3.2.9	Data Sampling .....	53

<b>CHAPTER 4 RESULTS AND DISCUSSIONS.....</b>	<b>55</b>
<b>4.1 Deep Hybrid Network (DHN) architecture .....</b>	<b>55</b>
<b>4.2 Experimental Results .....</b>	<b>58</b>
4.2.1 Results for 1kmx1km grids.....	58
4.2.2 Results for 5kmX5km grids.....	65
<b>4.3 Sensitivity analysis and feature importance .....</b>	<b>71</b>
4.3.1 Sensitivity analysis for spatial grids 1kmx1km .....	72
4.3.2 Sensitivity analysis for spatial grids 5kmx5km .....	73
<b>4.4 Comparison of different spatial resolution .....</b>	<b>74</b>
<b>4.5 Discussion summary .....</b>	<b>75</b>
<b>CHAPTER 5 CONCLUSION.....</b>	<b>77</b>
<b>5.1 Conclusions.....</b>	<b>77</b>
<b>5.2 Future Research Direction .....</b>	<b>78</b>
<b>REFERENCES.....</b>	<b>79</b>
<b>VITAE.....</b>	<b>90</b>

## LIST OF TABLES

Table 1: A comparison of different machine learning and deep learning models used in the literature .....	17
Table 2: Crash Attributes .....	32
Table 3: Hourly weather data description .....	33
Table 4: Intersection Red Light Violation .....	35
Table 5: Speed Camera Violation .....	38
Table 6: Taxi trips attributes description .....	39
Table 7: Bus ridership attributes description .....	41
Table 8: Official holidays of Chicago city in 2018.....	42
Table 9: Hyperparameter ranges and best hyperparameter.....	57
Table 10: Methods performance comparison .....	58
Table 11: Pairwise comparison of models performance using Tukey test for 1kmx1km spatial grids .....	64
Table 12: Performance of six models comparison.....	65
Table 13: Models performance result for November and October as tests months.....	67
Table 14: Pairwise comparison of models performance using Tukey test for 5kmx5km spatial grids .....	70

# LIST OF FIGURES

Figure 2: Chicago City Location in the US .....	4
Figure 3: Analysis Process.....	28
Figure 4: Data types to the model and expected output of the model.....	31
Figure 5: Intersection Red Light Camera Locations.....	35
Figure 6: Speed camera locations .....	37
Figure 7: Bus route in Chicago .....	41
Figure 8: Discretization of the study area into spatial grids .....	43
Figure 9: Crashes discretization into spatial grids .....	44
Figure 10: Bus stop density into spatial grids.....	44
Figure 11: Roads discretization into spatial grids.....	45
Figure 12: Intersection red light violations discretization into spatial grids.....	45
Figure 13: Road density discretization into spatial grids.....	46
Figure 14: Confusion Matrix .....	47
Figure 15: LSTM model architecture .....	51
Figure 16: DHN architecture .....	56
Figure 17: Comparison of AUC, recall, and accuracy for 1kmx1km spatial grids .....	59
Figure 18: Comparison of FAR for 1kmx1km spatial grids .....	59
Figure 19: ROC curve for all methods with 1kmx1km spatial grids.....	60
Figure 20: Spatial grid-level performance of methods in terms of AUC .....	61
Figure 21: Number of failed spatial grids in spatiotemporal crash prediction for all methods.....	62

Figure 22: Road density and crash frequency for each spatial grid .....	63
Figure 23: Comparison of AUC, recall, and accuracy for 5kmx5km spatial grids .....	66
Figure 24: Comparison of FAR for 5kmx5km spatial grids .....	67
Figure 25: ROC curve for all methods with 5kmx5km spatial grids .....	68
Figure 26: Performance of each spatial grid (5kmx5km) and number of crashes in the . test set (December) .....	69
Figure 27: Sensitivity analysis for 1kmx1km spatial grids.....	73
Figure 28: Sensitivity analysis for 5kmx5km spatial grids.....	74
Figure 29: Comparison of models in different spatial resolutions in terms of AUC .....	75

## **LIST OF ABBREVIATIONS**

1. AUROC = Area Under the Receiver Operating Characteristic
2. ANOVA = Analysis of Variance
3. ANN = Artificial Neural Network
4. BLR = Bayesian Logistics Regression
5. Bi-LSTM = Bidirectional Long Short-Term Memory
6. CLM = Conditional Logistic Model
7. CNN = Convolution Neural Network
8. DAP = Deep Accident Prediction
9. DHN = Deep Hybrid Network
10. DMLP = Deep Multi-Layer Perceptron
11. DNN = Deep Neural Network
12. EGB = Extreme Gradient Boosting
13. ERT = Extremely Randomized Trees
14. FAR = False Alarm Rate
15. FN = False Negative
16. FP = False Positive
17. FPR = False Positive Rate
18. GBT = Gradient Boosting Decision Tree
19. KNN = K-Nearest Neighbor
20. LSTM = Long Short-Term Memory
21. LSTM-CNN = Long Short-Term Memory Convolutional Neural Network

22. LR = Logistic Regression,
23. MAE = Mean absolute error
24. MARS = Multivariate Adaptive Regression Splines,
25. MLP = Multi-Layer Perceptron,
26. MSE = Mean square error
27. RF = Random Forest,
28. RMSE = Root mean square error
29. ROC = Receiver Operating Characteristic
30. SMOTE = Synthetic Minority Oversampling Technique
31. SVM = Support Vector Machines,
32. TNR = True Negative Rate
33. TA-STAN = Spatial-Temporal Attention Network
34. TN = True Negative
35. TP = True Positive

## ABSTRACT

Full Name : Mohammad Tamim Kashifi

Thesis Title : SHORT-TERM SPATIOTEMPORAL CRASH PREDICTION AND  
MODELING – APPLICATION OF DEEP LEARNING AND HYBRID  
MODELS

Major Field : Civil Engineering

Date of Degree : April/2021

**Traffic crashes are one of the significant causes of death worldwide. Proactive crash prevention can decrease crash risk. However, predicting road crashes in fine spatial resolution is complicated due to several contributing factors to a road crash and the crashes' random happening nature. This study predicts the spatiotemporal road crash using multi-source features, namely spatial, temporal, and spatiotemporal, collected to capture all possible crash triggering factors. Spatial features are road density, bus density, intersection, walkway, and bikeway density. Temporal features are weather conditions and events, and holidays. Similarly, spatiotemporal features include taxi trips, bus ridership, speed, and red-light violations. This study will adopt a wide variety of relevant machine learning models for spatiotemporal crash prediction. Initially, deep learning models like Long-term Short Memory (LSTM) and its combination with Convolution Neural Network (CNN) are considered to capture the natural complexity of the crash prediction. LSTM and CNN models' competitive advantage is due to their long-term memory and spatial recognition capacity, respectively. In this study, ANN, LSTM, and CNN models are combined as a hybrid deep learning model for spatiotemporal prediction of crashes in fine grids of Chicago. The main novelty of this study is to propose a deep hybrid network that uses spatial, temporal, and spatiotemporal jointly for crash prediction. Additionally, this study's deep hybrid network was compared with the state-of-art methods of the**

**literature. The traditional statistical logistic regression, artificial neural network, convolution neural network, long short-term memory method, and bidirectional LSTM are the baseline methods. The study results indicated that the deep hybrid network outperformed the baseline models in 1kmx1km and 5kmx5km spatial grids. The deep hybrid network achieved accuracy 0.722, recall 0.701, false alarm rate 0.278, and AUC 0.790 for spatial resolution of 1kmx1km and accuracy 0.781, recall 0.768, false alarm rate 0.216, and AUC 0.8586 for spatial resolution of 5kmx5km.**

## ملخص الرسالة

الاسم الكامل: محمد تميم كاشفي

عنوان الرسالة: التنبؤ والنمذجة لحوادث التصادم الزماني المكاني قصير المدى - تطبيق التعلم العميق والنماذج الهجينة

التخصص: الهندسة المدنية

تاريخ الدرجة العلمية: أبريل/2021

تعتبر حوادث المرور أحد الأسباب الرئيسية للوفاة في جميع أنحاء العالم. يمكن للوقاية الاستباقية من الاصطدام أن تقلل من مخاطر الاصطدام. ومع ذلك، فإن التنبؤ بحوادث الطرق بدقة مكانية عالية أمر معقد بسبب العديد من العوامل المساهمة في وقوع حوادث الطرق وطبيعة الحوادث العشوائية التي تحدث. هذه الدراسة تقوم بالتنبؤ بحصول حادث الطريق الزماني المكاني باستخدام خصائص وميزات متعددة المصادر، وهي المكانيّة، والزمانية، والزمانية المكانيّة، والتي تم جمعها للإمام بجميع العوامل المسببة للاصطدام المحتملة. الميزات المكانيّة هي كثافة الطريق، وكثافة الحافلات، والتقاطع، والممر، وكثافة طريق الدراجات. الميزات الزمنية هي أحوال الطقس، والأحداث والمناسبات، والعطلات. وبالمثل، تشمل الميزات الزمانية المكانيّة رحلات سيارات الأجرة، وركاب الحافلات، والسرعة، وانتهاكات الإشارة الأحمر. تتبنى هذه الدراسة مجموعة متنوعة من نماذج تعلم الآلة ذات الصلة بالتنبؤ المتعلق بحوادث التصادم الزماني والمكاني. بداية، تم اعتبار نماذج التعلم العميق مثل الذاكرة القصيرة طويلة المدى (LSTM) وبالجمع بينها وبين الشبكة العصبية الالتفافية (CNN) لالتقاط التعقيد الطبيعي للتنبؤ بالتحطم. ترجع الميزة التنافسية لنماذج LSTM و CNN إلى الذاكرة طويلة المدى وقدرة التعرف المكاني. في هذه الدراسة، تم دمج نماذج ANN و LSTM و CNN كنموذج هجين للتعلم العميق للتنبؤ الزماني والمكاني بالحوادث في الشبكات الدقيقة في شيكاغو. هذه الدراسة تقدم مقترحاً حديثاً في مجال البحث العلمي وهو شبكة هجينة عميقة تستخدم بشكل مشترك المكاني والزماني والمكاني الزماني للتنبؤ بحوادث الاصطدام. بالإضافة إلى ذلك، تمت مقارنة الشبكة الهجينة العميقة المقترحة في هذه الدراسة بأساليب ونماذج أخرى طرحت في الأبحاث العملية. حيث أن الأساليب والنماذج الأساسية هي الانحدار اللوجستي الإحصائي التقليدي، والشبكة العصبية الاصطناعية، والشبكة العصبية الالتفافية

، وطريقة الذاكرة طويلة المدى ، و LSTM ثنائي الاتجاه. أشارت نتائج الدراسة إلى أن الشبكة الهجينة العميقة تفوقت على النماذج الأساسية في كل من الشبكات المكانية 1 كم × 1 كم و 5 كم × 5 كم. أيضا، حققت الشبكة الهجينة العميقة دقة 0.722، واستدعاء 0.701، ومعدل إنذار خاطئ 0.278، ومنطقة تحت المنحنى 0.790 للدقة المكانية البالغة 1 كم × 1 كم. بالإضافة إلى الدقة 0.781، واستدعاء 0.768، ومعدل الإنذار الخاطئ 0.216، ومنطقة تحت المنحنى 0.8586 للدقة المكانية 5 كم × 5 كم.



# CHAPTER 1

## INTRODUCTION

With the rapid urbanization and economic development rate, there have been rapid automobile numbers worldwide. For instance, it is expected that around 85% of the United States Population will live in cities (United Nations, 2014). This is especially the concern in the United States, with the highest vehicle ownership per capita and the second-highest total vehicle after China in the world. The U.S. has a stock of 259.1 million vehicles and vehicle ownership of 832 vehicles/ 1000 persons as of 2016, and it has been continuously growing (Stacy & Robert, 2020). It resulted not only in chronic congestion in major cities but also in vast traffic crashes. For instance, Chicago downtown has experienced a 2.3% decrease in speed between 2016-2017 (Schaffer, 2019). Aside from chronic congestion, the rapid growth of urbanization and motorization leads to a traffic crash.

Road crashes the eighth leading cause of unnatural death in all age groups and the leading cause of unnatural death in children and young adults (age group 5-29 years) worldwide. In the U.S., it seems that road crash leads to a higher proportion of death than other causes of death than the rest of the world. Road crashes are the leading cause of unintentional death for people aged 1–54 in the U.S. (WISQARS, 2020).

Further, road crashes cause around 50 million injuries, leading to around 1.35 million fatalities yearly (WHO, 2018). In addition to the loss of life and property damage to the people involved in a crash, traffic crash puts much economic burden on the health care

system. Road injuries only cost \$518 billion worldwide annually on health care agencies (WHO, 2004). These are aside from property loss during a crash and traffic blockage and costs after a crash.

Timely traffic crash prediction can diminish the crash risk. The prediction would consider both the place where a crash would occur and the time to feed the Intelligent Transportation System (ITS) in proactive crash prevention. Proactive crash prevention can save human life and property. Traffic agencies can use crash prediction models to target crashes' specific times and locations to take measures proactively (Park et al., 2016). The model can be used for variable message signs (VMS) to provide real-time information for road users and improve road route choice for decreasing the probability of crashes (Yan & Wu, 2014). Additionally, dynamic spatiotemporal crash prediction can also be integrated with the emergency care system to feed real-time alarms on the possibility of a crash and provide emergency care accordingly.

However, traffic crashes are highly random and dynamic events, both considering the temporal and spatial occurrence. The involvement of many factors, like human characteristics, weather, infrastructure, and the vehicle, makes crash prediction highly sophisticated (Ren et al., 2018). Further, real-time crash-related data is challenging to achieve. For instance, human behavior is one of the leading causes of road crashes that varies between individuals, and it is not easy to acquire in real-time (Q. Chen et al., 2016). Spatiotemporal prediction of crashes in fine spatial resolutions is complicated. Although crash prediction in granular spatial resolution like county level and traffic zone level are discussed by some authors (Huang et al., 2010a; Rhee et al., 2016), fine spatial resolution crash prediction still seems is still not much addressed in the literature.

Considering the complicated nature of crash occurrence, some authors tried to collect big mobility-related data. For instance, Bao et al. (2018) used taxi trips Global Positioning System (GPS) density data as a factor of human mobility on-road crashes, and Chen et al. (2016) used cell phone GPS data as a human mobility factor. Although some authors use Twitter posts in real-time also as human mobility-related data, there is much noise in social media like spam posts (M. Chen et al., 2014).

Three types of data were collected: 1) spatial features, 2) spatiotemporal features, and 3) temporal features to predict traffic crashes. The spatial features vary based on the location of grides, but they are not changing over time. The spatial features are road density, bikeway density, bus stop density, and pedestrian route density. Spatiotemporal features vary both by time and space. Spatiotemporal features are intersection red light violation, speed violation, taxi trip traces, and bus ridership. Similarly, temporal features change by time consists of weather data and holidays and events data.

The study contributes to the literature in the following aspect. First, the proposed hybrid model in this study is a novel approach for crash prediction. To the best of our knowledge, such a method is not encountered in the literature for crash prediction. Further, the data types used in this study are inclusive of multiple sources. Twelve data sources were used for crash prediction. Further, a sensitivity analysis was conducted to explore the importance of each feature in model prediction.

## **1.1 Study Area**

Chicago is one of the largest cities in the United States, located on Lake Michigan in Illinois. Chicago has a total area of 606.1 km<sup>2</sup> and a metropolitan area of 10,874 km<sup>2</sup>. The population of the city in 2018 was 2.706 million (Data Commons, 2020). This city has the

highest population in the Midwestern United States. Chicago's total number of vehicles as of 2018 was 3,456,125 (Continental Motors, 2020). The vehicle ownership per household in 2015 was 1.11, increasing to 1.12 in 2016 (Maciag, 2017). Figure 2 indicates the study area of this study.

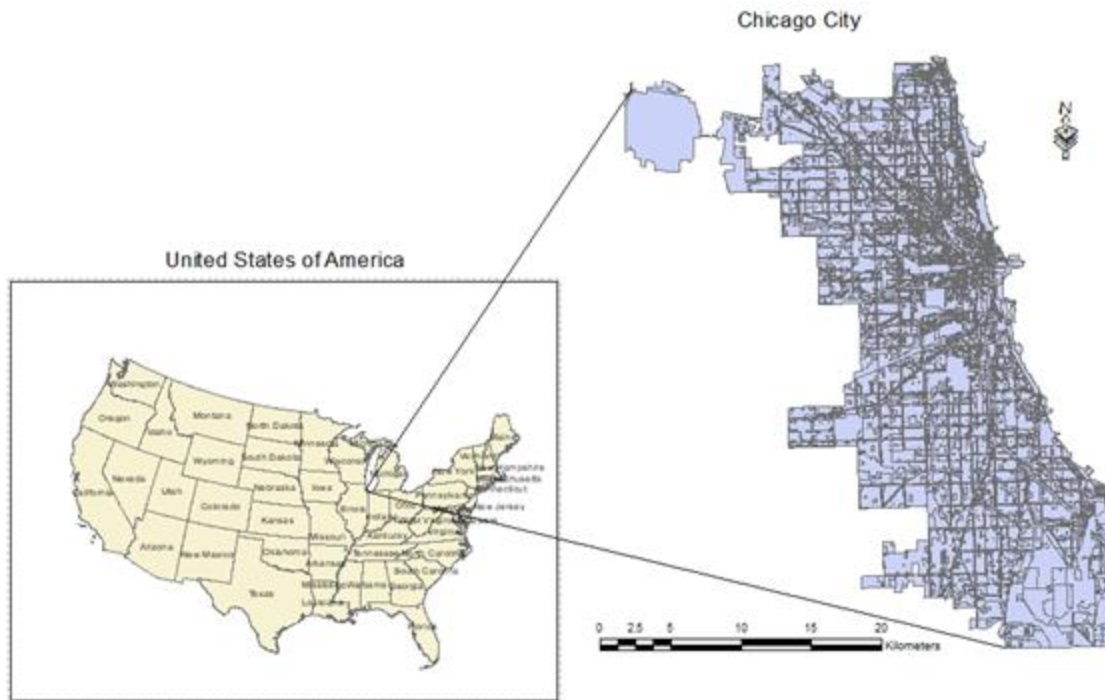


Figure 2: Chicago City Location in the US

## 1.2 Objectives

The objective of this study is multifold. The summary of the objectives as follows:

1. To investigate the influence of spatial, temporal, and spatiotemporal data on crash risk prediction for Chicago city.
2. To analyze and combine spatial, temporal, and spatiotemporal data for predicting crashes using hybrid deep learning models (like CNN-LSTM).

3. To develop a number of statistical and shallow machine learning models as a baseline to predict the crash risk for different levels of spatial aggregation starting from 1 km grid to 5 km grid.
4. To investigate the superiority of specific deep learning models and hybrid deep learning models to predict the crash risk for different spatial aggregation levels starting from 1 km grid to 5 km grid. |

### **1.3 Expected outcomes and their utility**

The expected outcome of this study would be multifold. First, The deep learning models were built and tested for spatiotemporal crash prediction, and the result of these models was compared with traditional statistical and machine learning models for comparison. These models would take historical crash data, real-time weather, and big data for representing human mobility like taxi trips for fine-resolution spatiotemporal crash prediction.

Spatiotemporal crash modeling and prediction is one of the components of smart cities. This component can be integrated into the Intelligent Transportation System (ITS) for real-time traffic management on the crash site. Further, crash prediction models can feed an advanced traveler information system to inform road users in advance of crash occurrence for route changing and traffic management after the crash occurrence. Timely prediction can assist emergency care agencies for real-time emergency services on the crash site.

Second, the influence of multi-source and big data were studied on crash occurrence and prediction. For instance, GPS density records are used in road crash modeling and

prediction. The importance of taxi trip data on crash modeling will be discussed. Further, the effect of a traffic violation on the real-time crash prediction will be studied.

## **1.4 Thesis Organization**

### **Chapter 1: Introduction**

In this chapter, the thesis introduction is presented. The chapter provides a comprehensive introduction to this thesis, including the data types, models, and methods. Further, the expected outcome, the study area, and the objectives of this study are presented in this chapter.

### **Chapter 2: Literature Review**

In this chapter, the related studies for crash prediction are discussed. The chapter introduces the types of methods are used for crash prediction. Then, the literature on road crash prediction on-road segments such as freeways and expressways are discussed. The studies regarding the spatiotemporal crash prediction on different levels of temporal and spatial resolution are discussed.

### **Chapter 3: Methodology**

The study methodology is presented in this chapter. The chapter presents the methodology of the thesis that starts with data acquisition and visualization and completes with the model test. The data types and descriptions are provided. Then the model's description is presented for five baseline models and the proposed hybrid model.

## **Chapter 4: Results and Discussions**

The results of the study are presented in chapter 4. The baseline models and the proposed models' performance were evaluated based on the four performance measures. Further, a sensitivity analysis was conducted to evaluate the significance of each data type. Then a significance test was conducted to explore the significant differences in model performance.

## **Chapter 3: Conclusion**

The study conclusions are presented in this chapter. The study's main conclusions are summarized, and the future direction of the research is also provided accordingly.

## CHAPTER 2

### LITERATURE REVIEW

In this chapter, an overview of existing literature pertinent to crash risk modeling and prediction is provided. The first section discusses methods of crash modeling and prediction. Then, crash spatial aggregation with different levels is described. After that, real-time crash prediction on different infrastructure elements is discussed in light of the literature.

#### 2.1 Crash Modelling and Prediction Methods

Predicting crash risk in real-time is generally considered by two methods, statistical models and machine learning methods (Yuan et al., 2019). Traditionally, statistical models are used for crash modeling and crash prediction using explanatory variables. These models are easy to interpret. For the real-time crash prediction on freeways by statistical models, matched case-control logistic regression (M. Abdel-Aty et al., 2004; C. Xu et al., 2012; Zuduo Zheng et al., 2010) and Bayesian model (M. A. Abdel-Aty et al., 2012) are frequently used to predict crash risk prediction in freeways.

However, statistical models relate the input parameters to the output classes by a particular model with specific assumptions (Shi & Abdel-Aty, 2015; Wang et al., 2015), leading to a lack of flexibility to adapt to complexity in crash prediction and modeling.

On the other hand, machine learning methods try to learn from the data with higher flexibility than statistical models (Alpaydin, 2020; Jordan & Mitchell, 2015; Marsland, 2015; Michie et al., 1994; Murphy, 2012) and can capture the non-linear and complex

relationships between features and output. Thus, machine learning models often result in better prediction capability than statistical models. Among machine learning models, Support Vector Machine (SVM), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and LSTM-CNN (Long Short-Term Memory Convolution Neural Network) are frequently used for real-time crash prediction on freeways, expressways, and intersections.

Nowadays, machine learning methods have been used for real-time predictions due to higher prediction capabilities. However, they are like a black box and lack interpretability and relationships. On the other hand, statistical models are well-known for their feature interpretation capabilities. The feature significance and importance determination are easy to understand by statical models. Thus, both models' combination is highly favored by many authors, e.g., (Shi & Abdel-Aty, 2015).

Recurrent, Neural Network (RNN), and Long Short-Term Memory (LSTM) are often used for real-time crash prediction. These models have the advantage of sequence learning, which is appropriate for time series analysis. In addition to capturing all other patterns and trends in data RNN and LSTM have memory and can capture the relationship between the seasonality of time series. Recurrent Neural Network (RNN) is a powerful model for sequence learning, such as natural language processing and time-series prediction (Tian et al., 2018; Zhao et al., 2017; Zibin Zheng et al., 2019). However, the RNN model has a gradient vanishing problem for long-term prediction and loses the prediction power in this case (Bengio et al., 1994). Long Short-Term Memory (LSTM) was invented in 1997, solves the long-term dependencies problem of RNN is especially useful for sequence learning (Hochreiter & Schmidhuber, 1997a).

## **2.2 Crash Spatial Aggregation for Modelling and Analysis**

The spatial aggregation analysis of crashes is of particular interest to researchers as safety is considered an essential component of transportation planning (Fischer et al., 2005; Washington et al., 2010). Proactive planning for the spatial units with high crash risk and taking measures can decrease the crash risk.

Aggregation of road crashes at different levels for analysis and prediction has been the interest of many researchers. For example, country-level aggregation (Noland, 2003), county-level crash aggregation and analysis (Aguero-Valverde & Jovanis, 2006; Amoros et al., 2003; Chang et al., 2011; Darwiche, 2009; Donaldson et al., 2006; Hanna et al., 2012; Huang et al., 2010b; Karlaftis & Tarko, 1998; Noland & Oh, 2004; Traynor, 2008), traffic analysis zone (TAZ) level (Bao et al., 2017; Rhee et al., 2016) level of crash aggregation has been analyzed by different authors to understand temporal and spatial patterns of crashes. Recently, grid-level aggregation, which is the discretization of a city or country into small square grids, say 5km x 5km, has attracted researchers' interest (Bao et al., 2019; Xie et al., 2017).

Huang et al. (2010) used the Bayesian spatial model to predict road crashes in Florida counties. The data used for this study includes five years of data (2003 to 2007) for 67 counties of Florida, including traffic crash data, traffic attributes, socioeconomic features, and location factors that reflect the proximity of areas. The author concluded that counties' crash frequency is positively correlated to traffic intensity, urbanization, and population density. In contrast, arterials and freeways crash frequency over Vehicle Mile Travelled (VMT), and the population is less intense.

Despite tremendous research on temporal and spatial crash analysis, the long-term nature of prediction, such as years or months, does not seem helpful for understanding real-time crash prediction and emergency resource allocation.

### **2.3 Crash Spatial Aggregation for Modelling and Analysis**

Real-time crash prediction requires real-time detectors data. The data are the speed of vehicles, traffic volume, the density of traffic. A traffic crash is predicted based on abrupt changes to traffic characteristics that are an abnormal situation. Abrupt changes to traffic characteristics are between successive traffic detectors that indicate an abnormal situation due to a traffic crash. Models detect these abnormal situations based on real-time data that is received from detectors.

Generally, the state-of-the-art detectors are classified into three types 1) intrusive detectors, 2) non-intrusive detectors, and 3) off-roadway technologies. The intrusive detectors, like inductive loops, are installed within the pavement. While non-intrusive, like microwave radar and video image processing, are installed above the road surface. Off-road technologies include probe vehicles and remote sensing. Intrusive detectors are commonly used types; however, they have some drawbacks. The intrusive detectors do not perform well when the road surface is in bad condition. Further, the installation and maintenance of intrusive detectors cause traffic disruption. Off-road technologies perform well for survey and data collection; however, they may not be the best option for real-time data collection as the real-time data needs the probe vehicle or similar technology to continuously run (Martin, 2003).

## **2.4 Real-time crash prediction on Freeways and Expressways**

Freeways crash prediction has been the interest of many researchers. This is due to the relatively stable traffic flow on freeways. Freeways flow is not much disturbed by intersections and other kinds of the junction. Thus, the occurrence of a crash disturbs the traffic flow and can be detected by abrupt changes to traffic characteristics upstream and downstream of the crash.

Shi and Abdel-Aty (2015) studied the real-time prediction of the operation and safety of Orlando, Florida expressways. The study's road network consisted of 75 miles of expressway with 275 Microwave Vehicle Detection systems (MVDS) that collected data flow in each lane based on a one-minute interval. The Random Forest (RF) was used to select features, and Bayesian models were used to predict rear-end road crashes from traffic flow dynamics and congestion index. Four features were significantly important (congestion index, log of traffic volume, peak hour, and mean speed of flow). Three types of Bayesian models, random effect, Fixed effect, and Random parameter, were used. The Bayesian with Random parameter outperformed the two other Bayesian types with a sensitivity of 68.8%, a specificity of 73.2%, accuracy of 72.2%, and AUC of 75.5.

Further, the author found that a wet road surface can increase crashes by 70%. The ramp metering technology and Dynamic Message Sign (DMS) were suggested to decrease the likelihood of expressway weaving segment crashes. The model's five performance measures are AUC of 70.1%, sensitivity 67.6%, specificity 70.0%, and accuracy of 69.8%.

Yu and Abdel-Aty (2013) used the Support Vector Machine (SVM) to predict real-time crash risk in a 15-mile mountainous freeway in Colorado. The feature selection for the models was based on the Classification And Regression Tree (CART) model. Then, the

author compared the real-time predictability of SVM with the conventional statistical model Bayesian logistic regression model. The performance measure used is the Area Under Curve (AUC). The author concluded that the SVM with a Radial basis kernel has an AUC of 77% that outperformed the SVM with a linear kernel with an AUC of 74% and Bayesian logistic regression with an AUC of 73%.

Parsa et al. (2019) implemented Probabilistic Neural Network (PNN) and Support Vector Machine (SVM) models to predict road crashes after 5 min of crash occurrence on the Eisenhower expressway in Chicago between June 2017 to December 2017.

Li et al. (2020) proposed LSTM-CNN (Long Short-Term Memory Convolution Neural Network) based crash risk prediction on eight-mile arterials. The model takes variables such as signal phasing, traffic flow parameters, and real-time weather data collected between September 2017 to September 2018 to predict crash risk in the next 5-10 min. The author also compared the model's result in five other models: XGBoost, LSTM, CNN, and Logistic Regression. The LSTM-CNN parallel model achieved AUC = 0.932 (with False alarm rate = 0.132 and sensitivity=0.868) and the five models achieved AUC less than 0.9 (Sequential LSTM-CNN AUC = 0.9, LSTM AUC= 0.88, CNN AUC=0.70, XGBoost AUC = 0.83, Bayesian Logistic Regression AUC = 0.7). The study concluded that the proposed LSTM-CNN model outperforms the other models in Area Under the Curve (AUC), false alarm rate, and sensitivity.

## **2.5 Real-time crash prediction on intersections and weaving sections**

A few studies also addressed crash prediction in real-time at intersections and weaving sections of freeways. For instance, Wang et al. (2015) used a multilevel Bayesian logistic regression model to predict real-time crashes on weaving segments of 22-mile expressways

with 16 weaving Florida segments in the next 5-10 min. The author argued that selecting 5-10 min before the crash data compared to 10-15 min produces higher accuracy than 0-5 min. It provides enough time for the ITS system to react. The data for the prediction of crashes are Microwave Vehicle Detection System (MVDS), geometry, and weather. The significant features for predictions were the main lane's speed at the start of the weaving segments, the logarithm of traffic volume, and the speed differences between the weaving segment's start and endpoint.

Yuan et al. (2019) used LSTM-CNN to predict real-time crash risk in 44 signalized intersections in Oviedo, Florida, in the next 5-10 min. Five datasets were used in this study which is 1) traffic crash database, which was collected between January 2017 to April 2018, 2) weather condition data, 3) loop detector data, 4) signal phasing dataset, and 5) travel speed data. The study excluded 2.35% of crash data due to the influence of alcohol and drug. Further, the author considered crashes within intersections (50.37% of total crashes) by stating that intersections are more predictable than outside intersections. The data split into 70% training, and 30% test and oversampling SMOTE was used to balance the dataset. The author compared the result of LSTM-CNN with the conditional logistic statistical model. The LSTM-CNN achieved a 60.67% sensitivity with a 39.33% false alarm rate, while the conditional logistic model got a 56.72% sensitivity with a 43.28% false alarm rate. The author declared that the model capability in prediction might not be practical due to low sensitivity and a high false alarm rate.

## **2.6 Machine Learning Method to Predict Real-time Spatiotemporal crash grid-based**

A traffic crash is a highly dynamic and random event. Short-term prediction of crashes is a highly challenging task. Similarly, predicting the crash location with a relatively fine spatial resolution like grids is highly complicated. Crash modeling for real-time prediction requires both the fine resolution of spatial grids and a short-term temporal timestamp for agencies to take action in advance of crash occurrence. Considering the challenging task of crash prediction with such leverages, the real-time crash prediction with the fine spatiotemporal resolution citywide or countrywide is still not studied (Bao et al., 2019).

Some studies attempted to predict road crashes at a spatial grid level for real-time prediction. For instance, Chen et al. (2016) used human mobility data to predict traffic crashes in real-time. The model used is deep Stack denoise Autoencoder and compared this model with a decision tree (DT), logistic regression (LR), and support vector machine (SVM). The data was 1.6 million users' GPS and seven-month traffic crash data (from January 01, 2013, to July 31, 2013) collected in Japan. The user mobility data was collected from mobile phones in 5 minutes. The study's temporal interval was considered one hour, and the spatial area was discretized as 500mx500m squares. The data was split as 80% for training and 20% for testing. The author concluded that the deep Stack Denoise Autoencoder model outperformed all other models in the study.

Ren et al. (2018) used both the deep learning model (LSTM) and the statistical model (e.g., ARMA) to predict spatiotemporal crashes in Beijing, China. The author discretized the city into a 1kmx1km area and the time into an hour. The data used for this study contained two years of the crash (2016 and 2017) with the timestamp and coordinates. The author

considered traffic crashes to predict traffic crashes without considering human mobility, calendar events, or traffic flow that would affect the model prediction capability. In the end, it was concluded that the LSTM model outperformed statistical models.

Bao et al. (2018) studied the effects of taxi trip patterns as a factor of human mobility on-road crashes. The author collected social-demographic attributes, taxi trips, GPS, road network data, land use data, and crash data. A total of 50 taxi trip patterns were found using the Latent Dirichlet Allocation (LDA) model. Geographical weighted Poisson regression (GWPR) was used to relate to road crashes and contributing factors. The modeling process compared the traditional contributing factors to road crashes and traditional factors and taxi trip patterns data. The author concluded that the taxi trip pattern data significantly improved the model prediction in the frequency of crashes. Other authors used the Global Position System (GPS) density of phone users as human mobility attributes to predict road crashes in Japan (Q. Chen et al., 2016).

**Table 1: A comparison of different machine learning and deep learning models used in the literature**

Study	Study Area	Timestamp	Models	Events	Data	Best Model	Variables	Performance Measure
(Basso et al., 2018)	Expressways	5 min	SVM,LR	Crash  Non-crash	39 crash  13,029 non-crash	SVM	type of vehicle, speeds, speed change, the variance of speeds, density, and density change	Accuracy = 67.89%  FPR=20.94%
(Li et al., 2020)	Arterials	5 min	LSTM-CNN, XGBoost,	Crash	432 crash	LSTM-CNN	traffic flow characteristics,	AUC=0.932

			BLR, LSTM,CNN	Non-crash	7,098,269 non-crash		signal timing, and weather conditions	FAR= 0.132  Sensitivity =0.868
(Yuan et al., 2019)	Signalized Intersections	5 min	LSTM-RNN, CLM	Crash  Non-crash	349 crash  3215 non-clash	LSTM-RNN	crash data, travel speed data, detectors installed, signal timing data , loop detector data ,weather characteristics	Recall=60.67%  FPR=39.3%

(Moosavi et al., 2019)	5kmx5km spatial grids	15 min	LR,GBT,DNN, DAP	Crash Non-crash	7,974 crash 29,020 non-crash	DAP	Crash data, weather, data, road data	Crash F1=0.65 Non-crash F1=0.89
(Zhu et al., 2019)	Taxi zones	12 hour	LR, LSTM, Xgboost ,TA-STAN	Crash Non-crash	Not given	TA-STAN	Taxi trips, city bike trips, weather data, crash data, street design data,	MSE = 0.000172 RMSE =0.01312 MAE = 0.0082

## **2.7 Machine Learning Methods to Predict Real-time Spatiotemporal crash grid-based**

Several authors discuss crash prediction on freeways and expressways; however, the real-time traffic data such as traffic volume, density, and speed are not available for the whole city or country road network. The sensors' data are mostly available on freeways, so the crash prediction by sensors data is limited to freeways and expressways based on the availability of sensors and Bluetooth data (Bao et al., 2019).

Further, installing and maintaining detectors requires massive investment for the whole city or country road network. Road sensors of any kind require regular maintenance. Considering the whole city road network, the sensors' installation and maintenance are a demanding task for road agencies. Thus, collecting sensors or Bluetooth data may not be economically feasible for all road networks and requires massive investment.

Fine-resolution spatial grid-based crash prediction is recently studied by some authors, e.g. (Bao et al., 2019; Xie et al., 2017). This method generally requires fewer data and predicts the road crash in each spatial grid based on the historical crash data itself, or we can add publicly available data like weather and taxi trips data.

However, the spatial grid-based crash modeling and prediction method are still not much explored in the literature. The existing literature mostly relies on LSTM or deep Stack denoise Autoencoder, e.g. (Bao et al., 2019; Xie et al., 2017). These models generally capture temporal relationships in the data and predict crashes accordingly. In contrast, the second dimension of relationships relies on the spatial associations of crash occurrence.

The short-term grid-based spatiotemporal crash prediction can be integrated into the Intelligent Transportation System (ITS) for real-time traffic management on the crash site. By this method, traffic agencies can manage traffic around the crash-prone area to avoid traffic crashes and study the spatial area for a possible reason for crash occurrence and solution to prevent a crash in the future.

Further, crash prediction models can feed an advanced traveler information system to inform road users in advance of crash occurrence for route changing and traffic management after the crash occurrence. Timely prediction can assist emergency care agencies for real-time emergency services on the crash site. Thus, crash severity can be decreased by providing timely medical support to the crash site.

## CHAPTER 3

### METHODOLOGY

In this chapter, the methodology of this study is described. The first part of the chapter deals with the problem setup, the study's mathematical modeling methods. In the subsequent sections, the analysis framework of the experiment is presented. This framework shows a step-by-step process of analysis which starts with data acquisition and ends with modeling testing.

#### 3.1 Problem Setup

In this section, the problem set up is described. Before any modeling, the inputs, outputs, and the purpose of modeling are determined. Further, the data need to be prepared into correct vectors and matrices for modeling.

1. The city is discretized into a vector of  $n$  grids  $g_n$  where  $G$  is the whole city represented as a vector of  $(n)$  elements.

$$\mathbf{G} = \langle g_1, g_2, g_3, \dots, g_n \rangle \in \mathbb{R}^N$$

The  $g$  is a spatial grid with a  $5\text{km} \times 5\text{ km}$  area as  $g_n \in G^N$ .

2. The time vector is a set of time intervals which is stamped at each  $t = 1$  hour for the study period  $T$  as:

$$\mathbf{T} = \langle t_1, t_2, t_3, \dots, t_k \rangle^T \in \mathbb{R}^{1 \times k}$$

Where each time interval is a step of the study period  $t_k \in T^{1 \times k}$ .

### 3.1.1 Inputs

As described in chapter 2, the explanatory features were classified into three types which are spatial features, temporal features, and spatiotemporal features. Each category of a feature needs to be mathematically prepared for machine learning models so that it be interpretable by models. Therefore, in this subsection, each category of features is mathematically formulated. The temporal and spatial features were presented as vectors as these categories of features are one-dimensional, while the spatiotemporal features were presented as matrices as they are two-dimensional, both spatial dimension and temporal dimension.

#### Spatiotemporal Features

Spatiotemporal features are described as matrices of  $n$  columns and  $k$  rows where  $n$  shows the gride position and  $k$  shows timestamp.

3. A dataset of taxi trips for  $n$  grides and  $t$  time intervals.

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{k1} & \cdots & x_{kn} \end{bmatrix} \in \mathbb{R}^{K \times N}$$

Where each taxi trip  $x_{tg}$  belongs to a spatial grid  $g_n$  and a timestamp  $t_k$  as

$$x_{tg} \in \{G^N, T^{1 \times K}\}$$

4. A dataset of speed violations for  $n$  grides and  $k$  time intervals.

$$\mathbf{S} = \begin{bmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{k1} & \cdots & s_{kn} \end{bmatrix} \in \mathbb{R}^{K \times N}$$

Where each speed violation event  $w_{tg}$  belongs to a spatial grid  $g_n$  and a timestamp  $t_k$  as

$$s_{tg} \in \{G^N, T^{1 \times K}\}$$

5. A dataset of intersection red-light violations for n grides and k time intervals.

$$\mathbf{L} = \begin{bmatrix} l_{11} & \cdots & l_{1n} \\ \vdots & \ddots & \vdots \\ l_{k1} & \cdots & l_{kn} \end{bmatrix} \in \mathbb{R}^{K \times N}$$

Where each red light violation event  $w_{tg}$  belongs to a spatial grid  $g_n$  and a timestamp

$t_k$  as

$$l_{tg} \in \{G^N, T^{1 \times K}\}$$

6. A dataset of bus ridership for n grides and k time intervals.

$$\mathbf{B} = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{k1} & \cdots & b_{kn} \end{bmatrix} \in \mathbb{R}^{K \times N}$$

Where each bus ridership  $b_{tg}$  the event belongs to a spatial grid  $g_n$  and a timestamp

$t_k$  as

$$b_{tg} \in \{G^N, T^{1 \times K}\}$$

## Temporal Features

Weather conditions and holidays and events temporal features. Temporal features were discretized as one hour.

7. The weather condition is described as a vector of  $W$  with k timestamps as

$$\mathbf{W} = \langle w_1, w_2, w_3, \dots, w_k \rangle^T \in \mathbb{R}^{1 \times K}$$

where each weather event  $w_k$  belongs a timestamp  $t_k$  as

$$w_k \in T^{1 \times k}.$$

8. The holidays and events are described as a vector of  $H$  with  $k$  timestamps as

$$\mathbf{H} = \langle h_1, h_2, h_3, \dots, h_k \rangle^T \in \mathbb{R}^{1 \times K}$$

where each weather event  $h_k$  belongs to a timestamp  $t_k$  as

$$h_k \in \mathbb{T}^{1 \times k}.$$

### **Spatial Features**

Spatial features vary only based on space, so they are represented as vectors.

9. The bus stop density feature is described as a vector of  $U$  with  $n$  elements for each spatial grids as

$$\mathbf{U} = \langle u_1, u_2, u_3, \dots, u_n \rangle \in \mathbb{R}^N$$

where each bus ridership  $u_n$  belongs to a spatial grid  $g_n$  as

$$u_n \in G^N.$$

10. The road density feature is described as a vector of  $R$  with  $n$  elements for each spatial grids as

$$\mathbf{R} = \langle r_1, r_2, r_3, \dots, r_n \rangle \in \mathbb{R}^N$$

where each element of road density  $r_n$  belongs to a spatial grid  $g_n$  as

$$r_n \in G^N.$$

11. The intersection density feature is described as a vector of  $I$  with  $n$  elements for each spatial grids as

$$\mathbf{I} = \langle i_1, i_2, i_3, \dots, i_n \rangle \in \mathbb{R}^N$$

where each element of intersection density  $i_n$  belongs to a spatial grid  $g_n$  as

$$i_n \in G^N.$$

12. The bikeway density feature is described as a vector of  $Y$  with  $n$  elements for each spatial grids as

$$Y = \langle y_1, y_2, y_3, \dots, y_n \rangle \in \mathbb{R}^N$$

where each element of bikeway density  $y_n$  belongs to a spatial grid  $g_n$  as

$$y_n \in G^N.$$

13. The pedestrian route density feature is described as a vector of  $P$  with  $n$  elements for each spatial grids as

$$P = \langle p_1, p_2, p_3, \dots, p_n \rangle \in \mathbb{R}^N$$

where each element of pedestrian route density  $p_n$  belongs to a spatial grid  $g_n$  as

$$p_n \in G^N.$$

### 3.1.2 Desired

14. A model  $M$  takes the features from  $(m)$  steps backward to predict each grid label in the next timestamp  $l_{(t+1)n}$ .

$$V = \{X, W\} \in \mathbb{R}^{K \times N}$$

$$V = \langle V_{t-m}, V_{t-m+1}, V_{t-m+2}, \dots, V_t \rangle^T \in \mathbb{R}^{1 \times K}$$

### 3.1.3 Output

15. The labels for each grid ( $g_n$ ) in the time step  $(t + 1)$  as  $l_{(t+1)n} \in \{0, 1\}$ .

$$L_{t+1} = \langle l_1, l_2, l_3, \dots, l_n \rangle \in \mathbb{R}^N$$

$$l_{(t+1)n} = \begin{cases} 1, & \text{if there is a crash in next hour} \\ 0, & \text{otherwise} \end{cases}$$

### **3.1.4 Purpose**

The purpose of the modeling is to minimize the model's prediction error or the mislabeled grides in the next timestamp. Therefore, different models, hyperparameter values, and data are used to minimize the classification error.

## **3.2 Analysis Framework**

The study analysis framework is presented in Figure 3. The process starts with data collection from multiple sources. Data discretization is required after data collection and preprocessing to make it readable for machine learning models. Data need to be split for training, validation, and test sets after the discretization of data. Data preprocessing and cleaning come in the fourth step. Then, it is needed to select performance measures for assessing model prediction capabilities. Models selection and training comes in the sixth step; training both the proposed and base models is necessary for comparison. After training, the models should be validated on the validation set of data for achieving optimum results. The model test is the last step of the process, which is conducted on the test set of data. The model performance is then reported based on the test set for model comparison. In the rest of this chapter, the framework is followed (Figure 3) and described each step of the framework in detail.

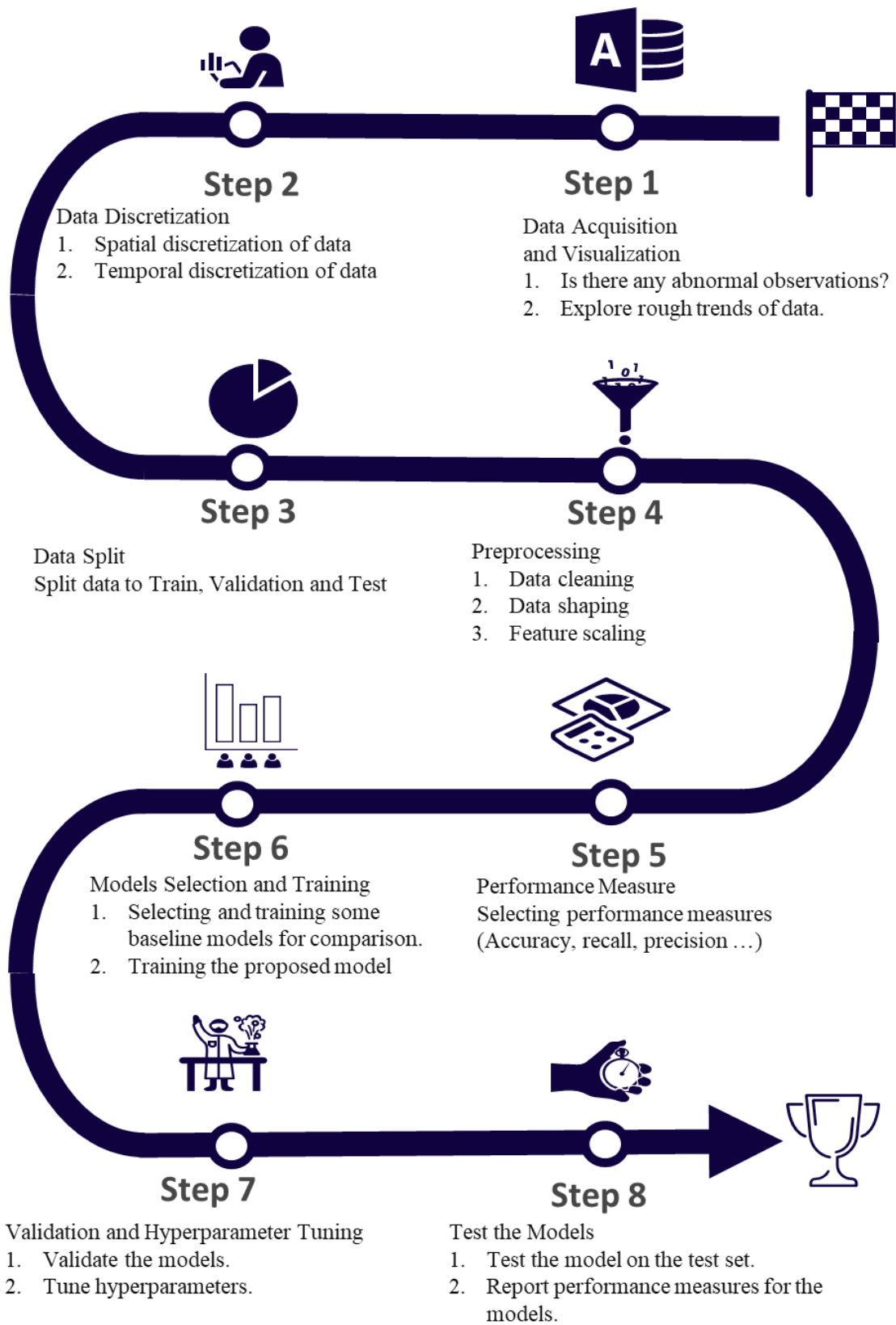


Figure 3: Analysis Process

### 3.2.1 Data

Several factors contribute to a road crash. These factors can be categorized into four major groups, which are 1) human error, 2) vehicle problem, 3) infrastructure, and 4) the weather condition. Therefore, several sources of data were collected for study to represent all possible contributing factors to crashes. Among them, taxi trips and bus ridership were used as human mobility factors. Some authors used social media data like Twitter data to represent human activity and mobility; however, there is much noise in social media like spam posts (M. Chen et al., 2014). Further, social media may show human activity, but it may not directly represent human mobility as taxi trips or bus ridership data. The following twelve types of data were used in this study;

1. Crashes dataset
2. Taxi trips database (a surrogate for human mobility and traffic exposure)
3. Holidays
4. Intersection red light violation
5. Speed violation
6. Weather condition
7. Road density
8. Intersection density
9. Pedestrian route density
10. Bikeway density
11. Bus ridership
12. Bust stop density

The data used in this research will be in three types, as shown in Figure 4. Three types of features, namely spatial, temporal, and spatiotemporal, were collected to capture all possible crash triggering factors. Spatial features are road density, bus density, intersection, walkway, and bikeway density; temporal features are weather conditions and events and

holidays. Spatiotemporal features include taxi trips, bus ridership, speed, and red-light violations.

Spatial features are those features that vary based on space. These features present local variations in crashes due to local variables. For instance, roads are an indicator of existing traffic that may lead to a traffic crash. Although there are several factors to crash occurrence, the higher the roads' density, the higher the road crash may happen. However, spatial features generally do not change over time. Thus, spatial features are used to capture spatial variations that lead to the crash occurrence.

Spatiotemporal features vary not only by space but also by time. For instance, taxi trips' GPS data indicates when the taxi trip happened, and also it has the coordinates where the taxi trip started and ended. These features are presented as matrices of time and space as described in the problem setup section. On the other hand, temporal features vary based only on time. For instance, public holidays are the same for the entire city. Weather condition data may vary by space; however, these variations are marginal, and the weather condition was considered the same for the entire city.

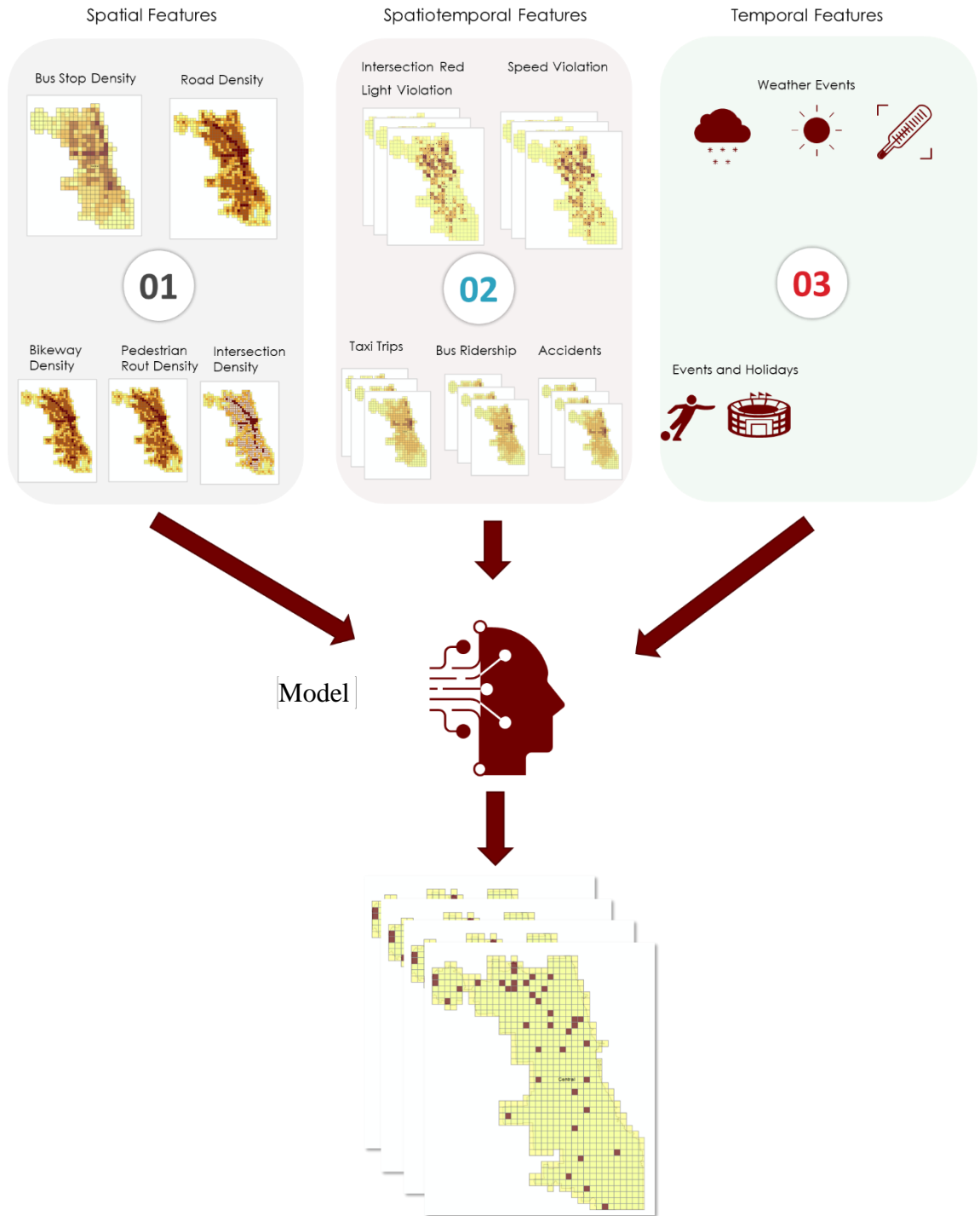


Figure 4: Data types to the model and expected output of the model

### 3.2.1.1 Crash Dataset

Traffic crash data was collected from the Chicago official public data portal (Chicago Data Portal, 2020). The crash data for the entire 2018 year (from January 01, 2018, to December 31, 2018) were considered. There were 118,079 crashes that happened during 2018. Crash attributes and their description is in Table 2.

**Table 2: Crash Attributes**

Column Name	Description	Type
CRASH_RECOR D_ID	This is a unique ID number for each crash.	Plain Text
CRASH_DATE_ EST_I	Crash date estimated by desk officer or reporting party (only used when the crash was reported after days of happening)	Plain Text
CRASH_DATE	The time of crash occurrence reported by the reporting officer	Date & Time
WEATHER_CO NDITION	Weather condition at the time of the crash. This feature has seven categories as 1) Clear, 2) Rain, 3) Snow, 4) Cloudy, 5) Fog, 6) Hail) and 7) Severe wind	Plain Text
LIGHTING_CO NDITION	Light condition at the time of the crash. This feature has five classes as 1) Daylight, 2) Darkness, Lighted Road, 3) Darkness, 4) Dusk, and 5) Dawn.	Plain Text
LANE_CNT	Total number of traffic lanes in the crash road (0 = intersection)	Number
ALIGNMENT	Alignment of the road where the crash occurred This feature has six classes as 1) Straight and level, 2) Straight on Grade, 3) Curve Level, 4) Straight on Hillcrest, 5) Curve on Grade, and 6) Curve on Hillcrest.	Plain Text
ROADWAY_SU RFACE_COND	Road surface condition. This feature has classes as five classes as 1) Dry, 2) Wet, 3) Snow or Slush, 4) Ice, and 5) Sand, Mud, or Dirt.	Plain Text
CRASH_TYPE	The severity of the crash. There are five levels of severity for each crash as No indication of injury: 101410 Non-incapacitating injury: 9247 Reported, not evident: 4964 Incapacitating injury: 2121 and Fatal: 114	Plain Text
LATITUDE	The latitude of the crash location	Number
LONGITUDE	The longitude of the crash location	Number

### 3.2.1.2 Weather Data

The weather condition data was also used as weather condition affects traffic crash occurrence. For instance, the driver behavior and response are different on a rainy or snowy day than a typical day. Thus, this difference may have an impact on crash severity and frequency (Mannering et al., 2016). Considering the effect of weather conditions on crash frequency and prediction, the weather data was considered as one type of temporal variable and incorporated in the models accordingly. Hence, weather data was received from Chicago Midway Airport (Arguez et al., 2020). The weather data are updated in the hourly rate with several attributes. Eight weather condition attributes related to a traffic crash were considered. The description of each weather condition attribute is given in Table 3. The weather-related features will further be studied to determine the importance of each feature on overall model performance quality improvement.

**Table 3: Hourly weather data description**

Column Name	Description	Type
Hourly Dry Bulb Temperature	The thermometer is measured in an open area, but the thermometer is protected from sunshine and moisture (°C).	Number
Hourly Precipitation:	Hourly precipitation (mm)	Number
Hourly Relative Humidity	The ratio of water vapor in the air at a given temperature (%)	Number
Hourly Sea Level	Standard sea-level pressure (mm)	Number

Pressure		
Hourly Sky Conditions	Cloud amount, height, and type	Plain Text
Hourly Visibility	Visibility reported by tower personnel (km)	Number
Hourly Wind Direction	The direction of the wind (deg)	Number
Hourly Wind Speed	Speed of Wind (km/h)	Number

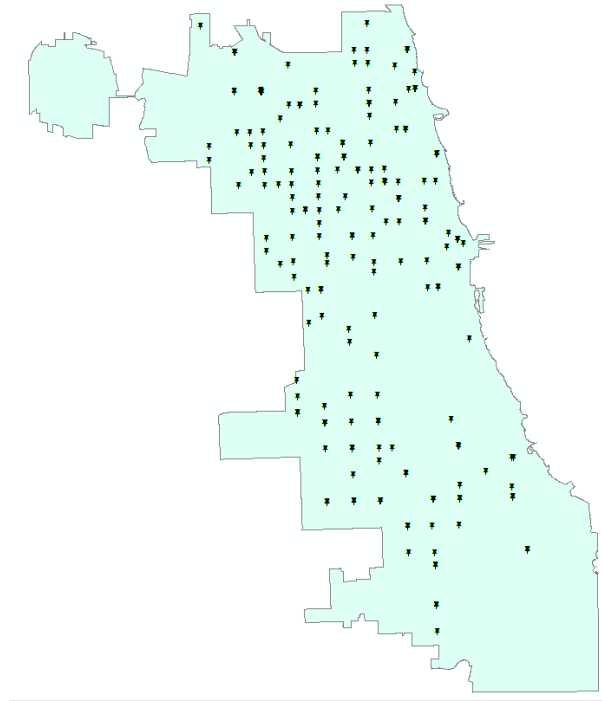
**3.2.1.3 Traffic Violations**

Traffic law violations are one of the contributing factors to a road crash. Thus the traffic law violations were considered in crash prediction for modeling. Chicago city is equipped with traffic enforcement law that enforces speed limit and intersection red light law throughout the city. In this study, traffic speed limit and intersection red-light violations represent the traffic violations' contributions to a traffic crash. The data is collected from the Chicago official public data portal (Chicago Data Portal, 2020).

***Intersection Red Light Violation***

There are 365 cameras located on intersections throughout the city of Chicago that enforce traffic red light law. Figure 5 shows the locations of inspection red light violation cameras. The red-light data are reported daily for each intersection. In this study, the effect of red-light violations on crash occurrence was considered. The location and day of violation

occurrence were both considered in the modeling process. The attributes of each violation in the intersections are described in Table 4.



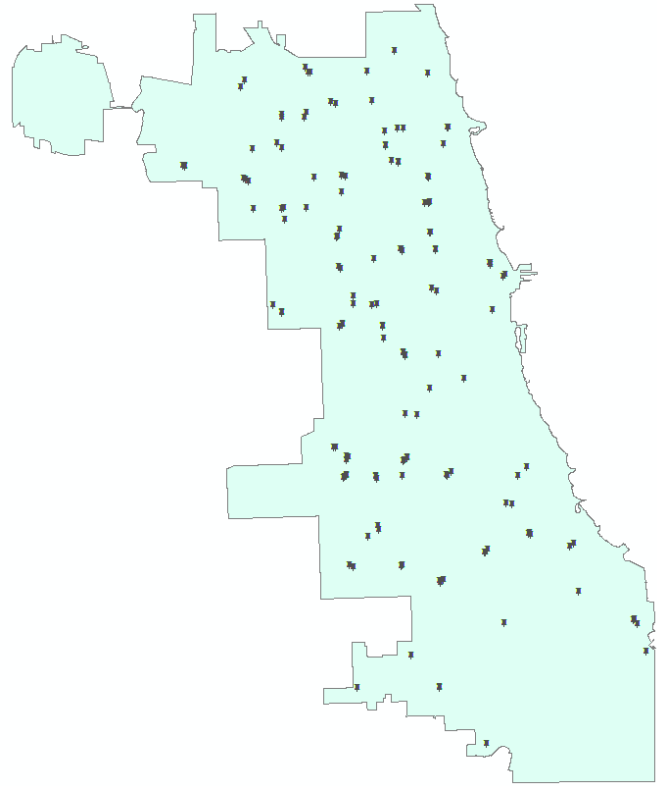
**Figure 5: Intersection Red Light Camera Locations**

**Table 4: Intersection Red Light Violation**

Column Name	Description	Type
INTERSECTION	The intersection where the red light camera is located for red-light law enforcement. Some intersections have more than one camera	Plain Text
CAMERA ID	The ID number of each physical red-light violation camera	Plain Text
ADDRESS	The address of each camera location in the physical intersection	Plain Text
VIOLATION DATE	The date of red-light violation occurrence.	Date & Time
VIOLATIONS	The number of red-light violations at each day	Number
LATITUDE	The latitude of camera location at a particular intersection. Geo-coded using the WGS84.	Number
LONGITUDE	The longitude of camera location at a particular intersection. Geo-coded using the WGS84.	Number

### ***Speed Limit Violation***

There are 162-speed limit cameras throughout Chicago city to enforce the speed limit, as shown in Figure 6. The violations in each camera are reported daily. The daily violations of each red-light enforcement camera alongside their locations to study their effect on the occurrence of a crash on each grid of the city were used. Attributes of each speed camera violation are described in Table 5.



**Figure 6: Speed camera locations**

**Table 5: Speed Camera Violation**

Column Name	Description	Type
ADDRESS	The physical location address of the speed enforcement camera(s) There may be more than one speed limit enforcement camera at each address	Plain Text
CAMERA ID	A unique ID for each physical camera at each location	Plain Text
VIOLATION DATE	The date of speed violation.	Date & Time
VIOLATIONS	Total number of speed violations at a particular camera per day	Number
LATITUDE	The latitude of the geographic location of speed camera location. Geo-coded using the WGS84.	Number
LONGITUDE	The longitude of the geographic location of speed camera location. Geo-coded using the WGS84.	Number

### **3.2.1.4 Taxi Trips**

Human needs movement to accomplish social and economic activities. This need for human mobility is the precursor of vehicle ridership that, in turn, leads to a traffic crash. Acquiring data of movement of each individual in a city is highly complicated to achieve due to the unavailability of devices for each individual movement to record, privacy concerns, and massive data analysis. Thus, some alternative data that can be achieved easily can be used as human mobility exposure. For example, taxi trips data that is recorded based on taxi trips pick up and drop off GPS location and their timestamps can be used as human mobility data. Taxi trips data is already available for some cities and can be readily acquired from taxi companies.

Bao et al. (2018) studied the effects of taxi trip patterns as a factor of human mobility on-road crashes. The author used Geographically Weighted Logistic Regression to explore the contribution of taxi trips data on road crash modeling and concluded that taxi trips' GPS data significantly improves the model for road crashes.

Taxi trips data in this study contained a total of 20,732,088 geocoded taxi trips with attributes as Table 6. The pickup and drop-off coordinates for each trip associated with each grid where they are located considered accordingly. The density of pick and drop of taxi trips was considered human mobility for each grid of this study area. Furthermore, the taxi pick and drop-off density may represent complete mobility as there are also other trip attributes like the trip duration and distance associated with each trip. Thus, the trip duration, distance of the trip, and total fare as other contributing factors associated with each trip as a mobility factor were used in this study.

**Table 6: Taxi trips attributes description**

Column Name	Description	Type
Trip Start Timestamp	The start time of the taxi trip rounded to the nearest 15 minutes for making the trip anonymous.	Date & Time
Trip End Timestamp	The end time of the taxi trip rounded to the nearest 15 minutes for making the trip anonymous.	Date & Time
Trip Duration in Seconds	Trip duration in seconds.	Number

Trip Kilometers Travelled	The distance of the trip in kilometers traveled.	Number
Fare Amount (USD)	The amount of fare for the trip.	Number
Tips (USD)	The tips, if there are any.	Number
Trip Total (USD)	The total trip cost.	Number
Pickup Centroid Latitude	The geographic latitude of the center of the pickup census tract	Number
Pickup Centroid Longitude	The geographic longitude of the center of the pickup census tract	Number
Drop-off Centroid Latitude	The geographic latitude of the center of the drop-off census tract	Number
Drop-off Centroid Longitude	The geographic longitude of the center of the drop-off census tract	Number

**3.2.1.5 Bus Ridership by Route**

In addition to taxi trip data, bus ridership was also considered as an indicator of human mobility. There are 127 bus routes in the entire city. The data of the bus route contains information regarding total bus ridership in each route daily. The bus route map of the city is shown in Figure 7, and other attributes of bus ridership data are presented in Table 7.

The bus ridership for each route was considered as a human mobility factor, and it was used in the crash prediction model.

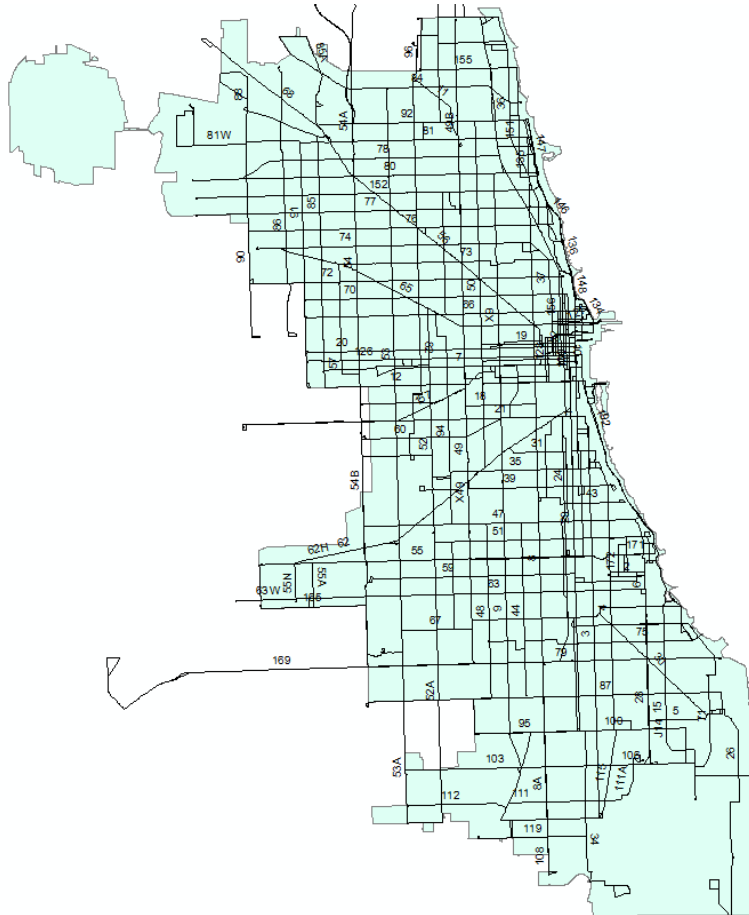


Figure 7: Bus route in Chicago

Table 7: Bus ridership attributes description

Column Name	Description	Type
route	The name of the bus route	Plain Text
date	Calendar data	Date & Time

day type	Indicates whether the day was a weekday, Saturday, or Sunday/holiday.	Plain Text
rides	Total number of rides in a particular route	Number

### 3.2.1.6 Holidays

Trip patterns during regular days and holidays differ. This affects the number of crashes that occur, also. Therefore, the dates of all holidays were considered an input to the models.

Table 8 shows the holidays and the dates of the holidays accordingly.

**Table 8: Official holidays of Chicago city in 2018**

holiday	date
New year' Day	1/1/2018
Martin Luther King Day	1/15/2018
Lincoln's Birthday	2/12/2018
President's Day	2/19/2018
Casimir Pulaski Day	3/5/2018
Mother's Day	5/13/2018
Memorial Day	5/28/2018
Father's Day	6/17/2018
Independence Day	7/4/2018
Labor Day	9/3/2018
Columbus Day	10/8/2018
Election Day	11/6/2018
Veterans Day	11/12/2018
Thanksgiving	11/22/2018
Day After Thanksgiving	11/23/2018
National Day of Mourning	12/5/2018
Christmas Eve	12/24/2018
Christmas Day	12/25/2018

### 3.2.2 Discretization to grids

Before analysis, the data need to be discretized so that machine learning models can understand the data. Machine learning models can understand numbers only. Temporal and spatial features should be represented in the vectors and matrices format as described in the problem setup section so that machine learning models understand the data. Thus, the Chicago city was discretized into grids. The grids of square shapes with different resolutions were considered. This study will consider experimenting with grids from 1kmx1km to 5kmx5km and see the model prediction based on different resolution scales. Further, the time was discretized into a one-hour interval. Figure 8 to Figure 13 shows the discretization of different features for this study into spatial grids.

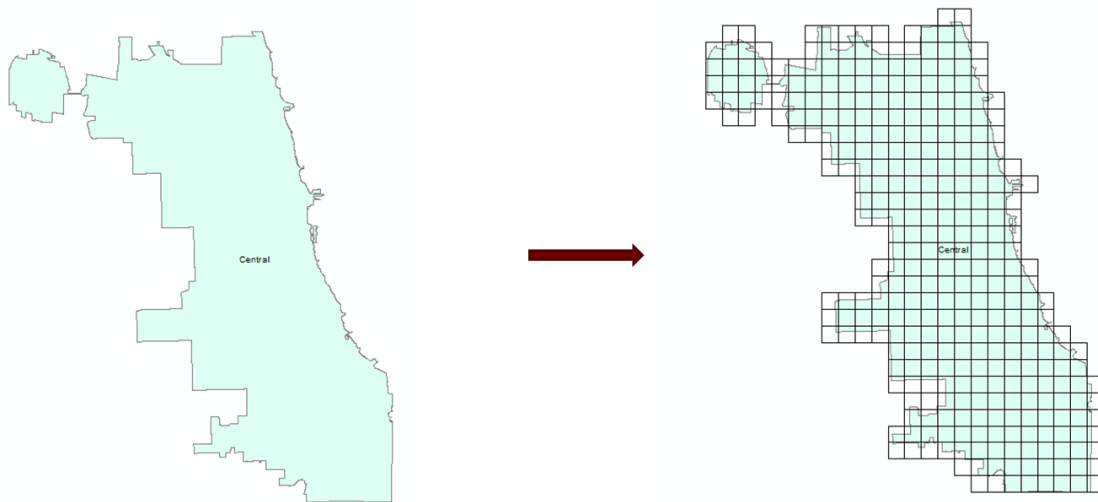


Figure 8: Discretization of the study area into spatial grids

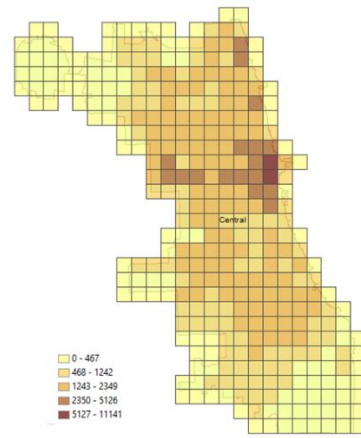
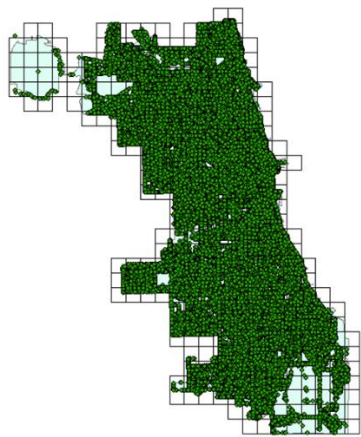


Figure 9: Crashes discretization into spatial grids

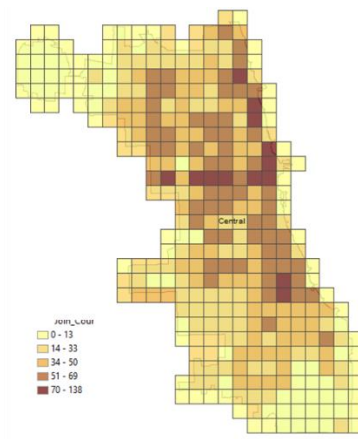
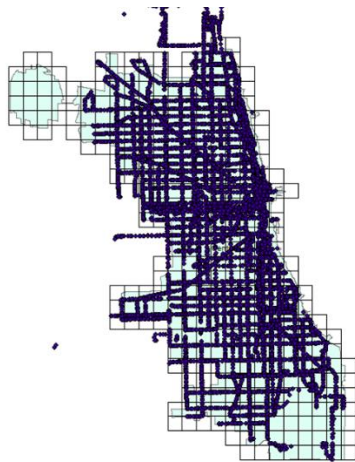


Figure 10: Bus stop density into spatial grids

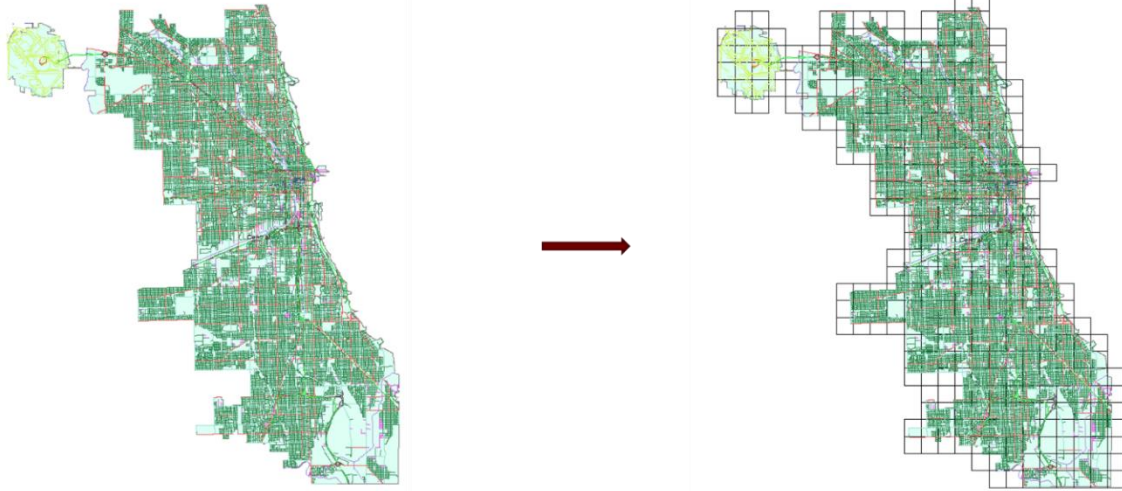


Figure 11: Roads discretization into spatial grids

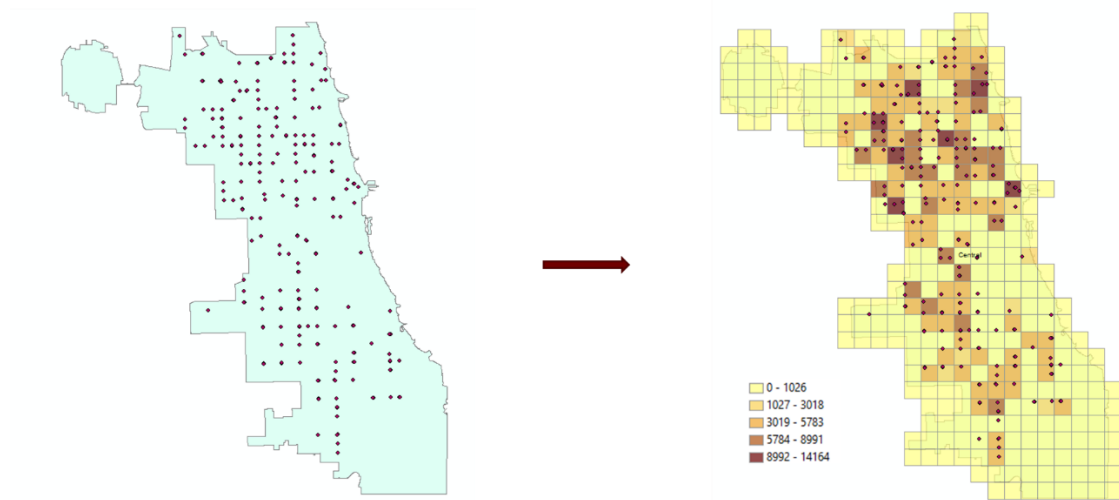


Figure 12: Intersection red light violations discretization into spatial grids

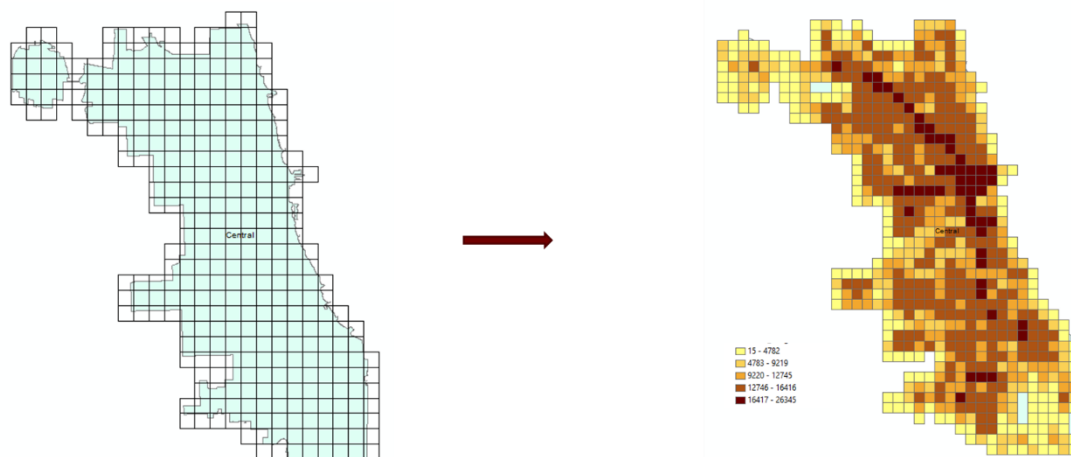


Figure 13: Road density discretization into spatial grids

### 3.2.3 Data Split

For machine learning modeling, data need to be split for training, validation, and test. The training set is used to train the model, and the validation set is used for hyperparameter tuning. Similarly, the test set is used for testing the model prediction capability based on untouched data that is not used for training and hyperparameter tuning. This is highly important to process each set of data separately to avoid model prediction bias. The data for the whole of 2018 were considered for this study. The first ten months' data were used for model training (January – October). The November month data is used for validation purposes, and December data was used for the model test.

### 3.2.4 Data Preprocessing

Before modeling with data, the data needs to be preprocessed. Feature scaling, for instance, is essential to decrease the model convergence time and training. Thus, the min-max scaling was used to scale the data before modeling. Further, some categorical data like the day of the week needs to be one-hot encoded for modeling. One hot encoding of features is conducted by showing the presence and absence of categorical data by 1 and 0,

respectively. Additionally, the data dimension needs to be configured for models that require a sequence of past data for future prediction. Therefore, the data shaped into three-dimensional matrices such as the first dimension shows the number of samples, the second dimension shows the window size, and the third dimension shows the feature numbers for models LSTM, Bi-LSTM, CNN, and DHN.

### 3.2.5 Performance Measures

In this study, four different performance measures are used. Accuracy, recall, false alarm rate (FAR), and area under Receiver Operator Characteristic (ROC) curve are used as performance measures in this study. To better understand these performance measures, a confusion matrix (Stehman, 1997) is used. In this matrix (Figure 14), the rows and columns are divided into actual and predicted classes. True positive (TP) cases are those cases that are positive and also predicted as positive by a classifier. Similarly, true negative (TN) is the negative cases that are predicted as negative. However, the false-negative (FN) and false positive (FP) are misclassified cases by a classifier. False-negative are the cases that are predicted as a negative class while they belong to the positive class. False-positive are those cases that are predicted as positive while they are from a negative class.

	<i>Predicted Positive</i>	<i>Predicted Negative</i>
<i>Actual Positive</i>	TP	FN
<i>Actual Negative</i>	FP	TN

Figure 14: Confusion Matrix

Accuracy is mainly used as a measure of choice in classification problems. The accuracy is the rate of total true predictions over the total sample size. Recall or sensitivity is the prediction power of a model. This measure shows the rate that a positive class is predicted as positive in a model. The FAR is the rate of non-crash events that were classified as a crash.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Recall (Sensitivity) = \frac{TP}{TP+FN}$$

$$FAR = \frac{FP}{FP + TN}$$

Receiver Operator Characteristic (ROC), which shows a trade-off between recall and FAR, is used to understand the prediction of a model under various rates of recall and FAR. Area Under Curve (AUC) is the area under the ROC curve, which ranges between 0.5 to 1. It is a summary statistic that shows the overall model performance overestimates the model performance for imbalanced data (Ozenne et al., 2015).

### **3.2.6 Model Selection and Training**

In this study, five baseline models are Logistic Regression, Artificial Neural Network, Convolution Neural Network, Long Short-term Memory, and bidirectional LSTM. Additionally, a hybrid model that combines deep learning models is proposed. The Python language was used to code the models and to compare the results. In the Python language, the Keras library was used, a state-of-the-art library for deep learning models with high

flexibility and adaptability. Further, for displaying the models' results on maps, the ArcGIS software was used.

### 3.2.6.1 Logistic Regression

Logistic regression is one of the commonly used linear statistical classifiers (Train, 2003). This classifier maximizes the likelihood of classification probabilities on training data. The model fits a series of hyperplanes on training data to separate between classes of labels. To avoid overfitting, the regularization hyperparameter is used. This hyperparameter controls the contributions each explanatory variable should have on the model output.

$$P(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_i X_i)}}$$

Where  $\beta_0$  is the model intercept,  $\beta_i$  is the weight of the explanatory variable  $X_i$ , and  $P(x)$  is the likelihood. The logit function or the link function of the logistic regression is as follows:

$$\text{logit}[p(x)] = \log\left[\frac{p(x)}{1 - p(x)}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

### 3.2.6.2 Artificial Neural Network (ANN)

ANN is made of input, hidden, and output layers. The input and output layers determine the input parameters dimension plus a bias and the number of output channels according to the problem, respectively. Each hidden layer processes information and assigns a weight to each neuron by a weight matrix that is learned during the training model (Hochreiter & Schmidhuber, 1997b). The learning process of ANN is conducted through a supervised backpropagation, and a decent gradient algorithm is used to minimize training loss in each iteration. The output of each neuron is then activated by an activation function. Activation

functions such as sigmoid, hyperbolic tangent, Rectified Linear Function ReLU (Gers et al., 2000).

### 3.2.6.3 Long-short Term Memory (LSTM) and Bidirectional LSTM

The LSTM model was proposed by (Hochreiter & Schmidhuber, 1997c). This model has the capability of long-term memory, and it is much used for sequence learning. Thus, the model has a particular application in natural language processing analysis and time-series analysis. Daily, around 4.5 billion translations are made by Facebook using LSTM (Olah, 2015).

Each unit of LSTM is connected to the following unit as a chain and transfers information to each other in sequence. In each unit of LSTM, four gates regulate the information transfer between nodes. The forget gate  $f_t$ , for instance, controls the amount of information from the previous cell state  $c_{t-1}$  which is represented by the hidden state  $h_{t-1}$  to be transferred to the current cell state  $c_t$ . The forget gate output is in the form of a sigmoid function whose outputs between 0 and 1 with one means all information to be transferred to cell state, and 0 means no information to be transferred. The input gate  $i_t$  controls the amount of information from features at the time  $t$   $x_t$  to be sustained in the cell state. The output gate  $o_t$  decides the amount of information from the hidden state  $h_{t-1}$  to be transferred to the next unit hidden state  $h_t$ . The  $a_t$  is a vector of candidate values used along with the input gate and forget gate to update the cell state (X. Zheng & Chen, 2021). The architecture of the LSTM (Liu & Guo, 2019; X. Zheng & Chen, 2021) model is shown in Figure 15.

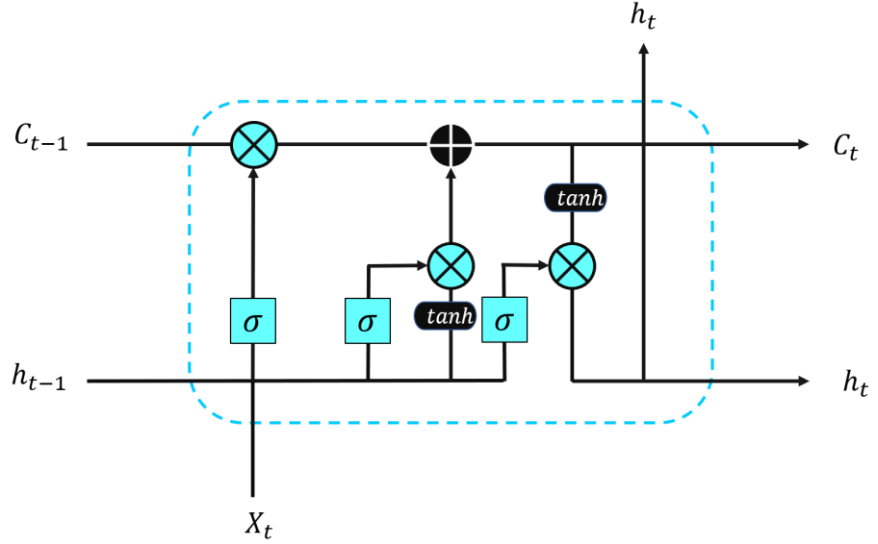


Figure 15: LSTM model architecture |

$$f_t = \delta(X_f x_t + H_f h_{t-1} + C_f c_{t-1} + b_f)$$

$$i_t = \delta(X_i x_t + H_i h_{t-1} + C_i c_{t-1} + b_i)$$

$$o_t = \delta(X_o x_t + H_o h_{t-1} + C_o c_{t-1} + b_o)$$

$$a_t = \delta(X_a x_t + H_a h_{t-1} + C_a c_{t-1} + b_a)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot a_t$$

$$h_t = o_t \odot \tanh(c_t)$$

In the above equations,  $\odot$  represents the element-wise multiplication operator, X, H, and C are the weights, b are biases.

Another variation of the LSTM model is bi-directional LSTM. In bi-directional LSTM, the hidden state is represented as Eq. 1 (Lecun et al., 2015)

$$h_t = \delta(W_h [h_t^{\rightarrow}, h_t^{\leftarrow}] + b_h) \quad 1$$

The bi-directional LSTM model has an advantage over the traditional unidirectional LSTM by capturing both previous and next windows in time series. The bi-directional LSTM applies forward and backward LSTM layers for data sequence long-term dependencies capturing (LeCun et al., 1989).

#### **3.2.6.4 Convolutional Neural Networks**

Convolutional Neural Networks (CNN) is another particular type of feed-forward neural network commonly used for visual analysis. The data for CNN models are generally in the form of multiple arrays like 1D arrays for signals analysis, 2D for images, and 3D volumetric image analysis. The CNN model is generally composed of convolution and pooling layers with the dense layer as the last layer. The convolution layer detects local correlations of features between successive layers, while the pooling layer's role is to combine similar features. The dense layer in CNN architecture aligns the feature's dimension from the pooling layer to the output classes (Núñez et al., 2018). CNN is highly effective in determining the data's spatial patterns (D. Xu et al., 2020).

#### **3.2.6.5 Hybrid Deep Learning Model**

CNN-LSTM is a combination of the CNN and LSTM models. This model attracted the interest of researchers in recent years. This model's flavor over CNN and LSTM is that this model captures both spatial and temporal patterns. In recent years, hybrid models applied in different fields to analyze sequence learning. The hybrid model's spatiotemporal prediction capability is reported to outperform LSTM and CNN models in many studies. For instance, (Núñez et al., 2018) used CNN-LSTM to predict the skeleton-based activity and gesture of hands classification, and the model outperformed the LSTM model.

### **3.2.7 Validation and Hyperparameter Tuning**

Model validation and hyperparameter tuning are conducted after training models. In this step, the hyperparameters for each model are tuned to explore the best hyperparameters. There are some already built optimization algorithms. Among them, Bayesian algorithms are commonly used. Thus, Bayesian hyperparameter tuning was used for hyperparameter tuning to find the best hyperparameters for each model.

The Bayesian model optimization is better than the random search and grid search as it considers the result of past trial performance and takes an informed decision on the selection of the subsequent values of hyperparameters. On the other hand, the random search and grid search hyperparameter tuning does not consider the past trials' performance in selecting the values of the hyperparameters in the subsequent trial. Therefore, the Bayesian optimization reaches better model performance in a fewer number of hyperparameter tuning. The Hyperopt library of Python, an already built model for Bayesian hyperparameter tuning, was used in this study.

### **3.2.8 Testing the models**

In the last step of modeling, models' prediction needs to be tested on the test set. The December month data was used for testing the models. It is important to note that the training set data should not be touched until the completion of already mentioned modeling steps to protect the models against the biased result. The performance measure of each model should be reported accordingly, and the models should be compared to each other.

### **3.2.9 Data Sampling**

There are generally two methods for data sampling to overcome the data imbalance problem: undersampling and oversampling. However, these methods are designed only for one output problem (D. Xu et al., 2020). This study problem is a multi-output problem. In

this study, the loss function was modified to compensate for the imbalanced data issue. The modified loss function is considered a weight for crash events that is equal to the number of non-crash events divided by crash events. This weight modifies the traditional loss function considers and impacts the learning process.

## CHAPTER 4

### RESULTS AND DISCUSSIONS

In this chapter, the results and discussions of the study are presented. After training and validating the proposed deep hybrid network (DHN) and the baseline models, the models were tested on different datasets (December 2018 data). The results of analysis for each model were reported and compared accordingly. Additionally, a sensitivity analysis was conducted to explore features' importance on model prediction.

#### 4.1 Deep Hybrid Network (DHN) architecture

In this study, a deep hybrid architecture was proposed that is constructed by connecting three types of neural networks (1. ANN, 2. LSTM, and 3. CNN). As shown in Figure 16, the DHN architecture consists of three branches. There is a temporal branch that inputs the temporal features, and the model applies the LSTM layers and dropout. The LSTM layers are very successful in the temporal analysis as it remembers the long-term temporal dependencies. The second branch of DHN takes the spatiotemporal features and applies the LSTM and CNN layers, respectively. The spatiotemporal features, as described earlier, varies based on space and time; thus, LSTM was applied to capture the temporal dependencies, and the CNN layer was applied to capture the spatial dependencies. The third branch of DHN is an embedding layer that takes the spatial variables and embeds the spatial variances into the model. The three branches of the DHN are merged, and two dense layers were applied at the end. The sigmoid function was used as the output layer that produces the probability of crash occurrence in each spatial grid in the next timestep. Then,

the probabilities are converted to the dichotomous outputs as zero and one with zero no crash, and one as a crash prediction.

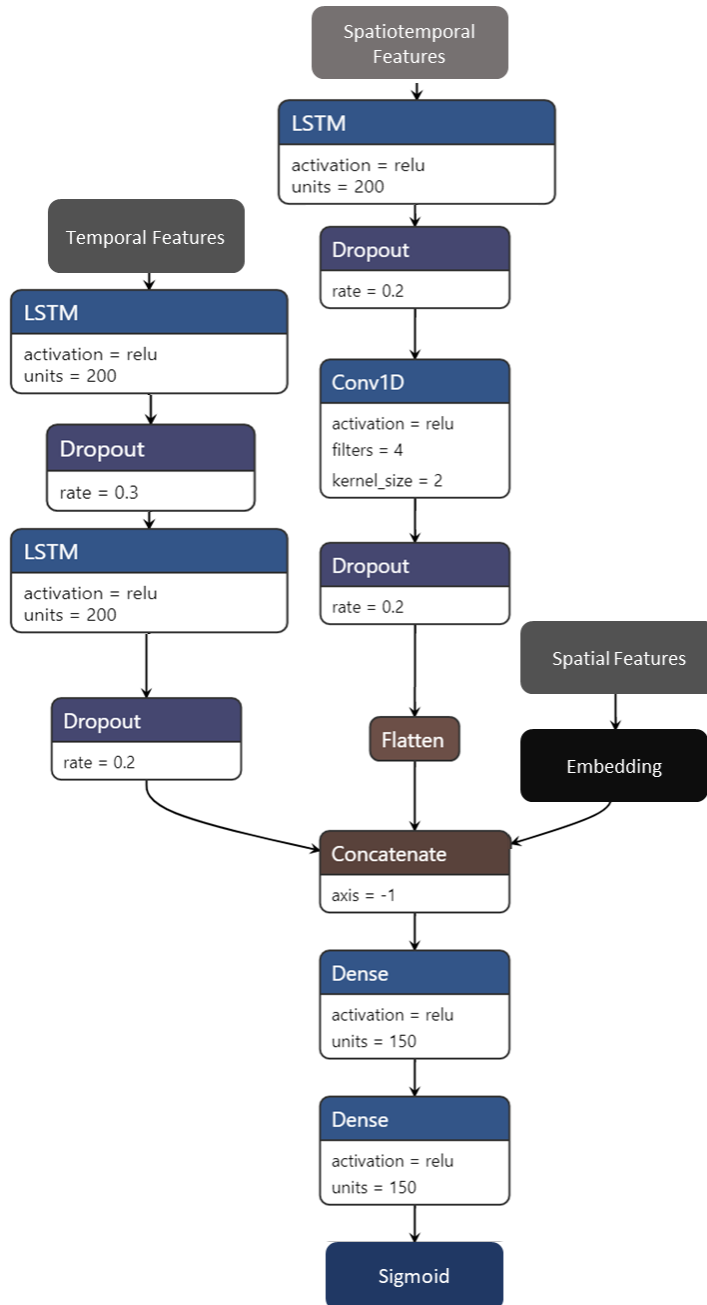


Figure 16: DHN architecture

The hyperparameter tuning was conducted iteratively to find the best hyperparameter for each model. The best hyperparameters were found based on hyperparameter tuning on the

validation data set (November 2018 data). The validation hyperparameters range the best hyperparameters are shown in Table 9. As can be seen in Table 9, the best values of dropout hyperparameters are 0.2 and 0.3 in two different branches of DHN. The lower values of dropout rate lead to models overfitting, while the higher values of dropout rate will result in the model underfitting.

Similarly, the number of dense layer neurons' optimum number is 150 neurons. This value of the number of neurons was obtained after hyperparameter tuning in the range of 50-500 neurons. Further, the optimum number of units in the LSTM layer was obtained as 200 units. Meanwhile, the CNN filters, CNN kernel size, and activation function were optimized in hyperparameter tuning. The result of these hyperparameter tunings is indicated in Table 9.

**Table 9: Hyperparameter ranges and best hyperparameter**

<b>No</b>	<b>Hyperparameter</b>	<b>Range</b>	<b>Best Hyperparameter</b>
<b>1</b>	Dropout rate	0.1-0.5	0.2,0.3
<b>2</b>	Dense layer neurons	50-500	150
<b>3</b>	LSTM units	50-300	200
<b>4</b>	CNN filters	2-132	4
<b>5</b>	CNN kernel size	2-8	2
<b>6</b>	Activation function	Sigmoid, Tanh, ReLU	ReLU

## 4.2 Experimental Results

The experimental results are generated for the test set of data after model training and validation. The experimental results are for the one-month test set (December of 2018). The results are presented for two spatial resolutions (1kmx1km and 5kmx5km grids).

### 4.2.1 Results for 1kmx1km grids

The results for 1kmx1km spatial grids are presented in Table 10. The performance measure reported in this study is AUC, accuracy, recall and false alarm rate (FAR). The results indicated that the LR has the lowest performance quality than the rest of the methods in all of the performance measures. The ANN, CNN, LSTM, and Bi-LSTM show similar performance without significant changes in their performances. However, the DHN method outperforms the rest of the methods in terms of AUC, accuracy, and FAR.

Table 10: Methods performance comparison

classifiers	AUC	Accuracy	Recall	False Alarm Rate	Training Time (s)
LR	0.7636	0.6836	0.6952	0.3167	113
ANN	0.7822	0.7043	0.7014	0.2956	187
CNN	0.7819	0.7139	0.6940	0.2857	195
LSTM	0.7829	0.7122	<b>0.7025</b>	0.2876	294
Bi-LSTM	0.7845	0.7167	0.6977	0.2830	319
DHN	<b>0.7898</b>	<b>0.7221</b>	0.7011	<b>0.2775</b>	<b>251</b>

Figure 17 indicates the AUC, recall, and accuracy for the six methods in this study.

According to the figure, the deep learning models significantly outperformed the

statistical model (LR). The LR has an AUC of 0.7636, which is 12% less than the rest of the methods. The highest AUC is 0.7898, which belongs to DHN.

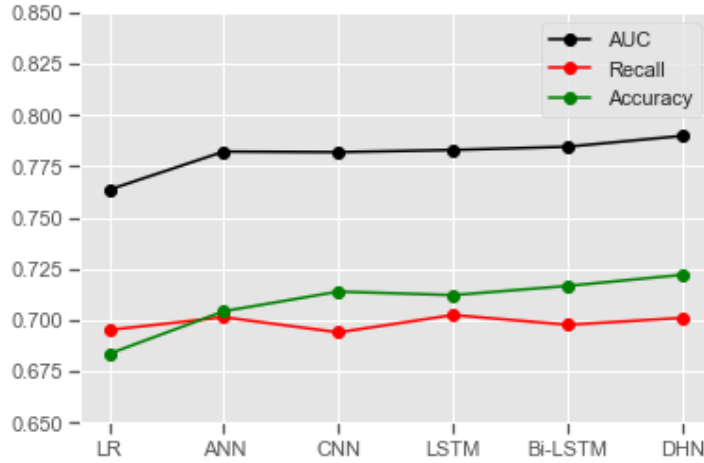


Figure 17: Comparison of AUC, recall, and accuracy for 1kmx1km spatial grids

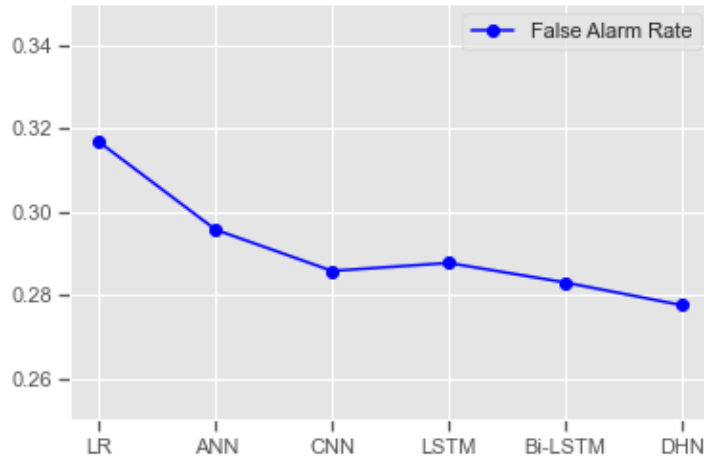
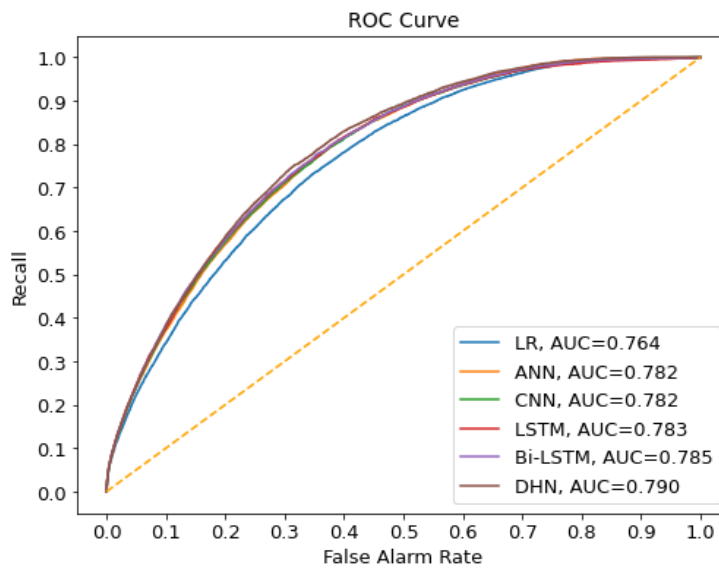


Figure 18: Comparison of FAR for 1kmx1km spatial grids

Similarly, Figure 18 shows the FAR of the six methods. The LR has the highest FAR, which is 0.3167, showing the lowest performance among the methods, while the DHN has

the lowest FAR, which is 0.2775. The rest of the deep learning methods have a similar FAR that is between LR and DHN in terms of FAR performance.

For a better understanding, the trade-off between recall and FAR, the ROC was plotted for all methods (Figure 19). This figure relates the recall and FAR. The acceptable FAR and their corresponding recall can be obtained accordingly. The policymakers should decide on the trade-off between FAR and recall that is suitable to the agency.



**Figure 19: ROC curve for all methods with 1kmx1km spatial grids**

The spatial heterogeneity of models' performance for each spatial grid of Chicago is indicated in (Figure 20). The criteria of each spatial grid performance in light of road density and crash frequency at spatial grid level (Figure 22) needs to be further investigated to understand the spatial distribution of model performance. For the binary classification task, the AUC value for random guess is 0.5, so any value below AUC below 0.5 means the model performance is worse than a random guess, and the model is not applicable. Therefore, spatial grids with AUC below 0.5 are considered as failed grids. Considering

these characteristics, the number of failed grids for each spatial grid are counted and compared accordingly. For LR, the spatial grids with AUC below 0.5 are distributed all over the city except the central business district. The rest of the models have a fewer number of failed spatial grids.

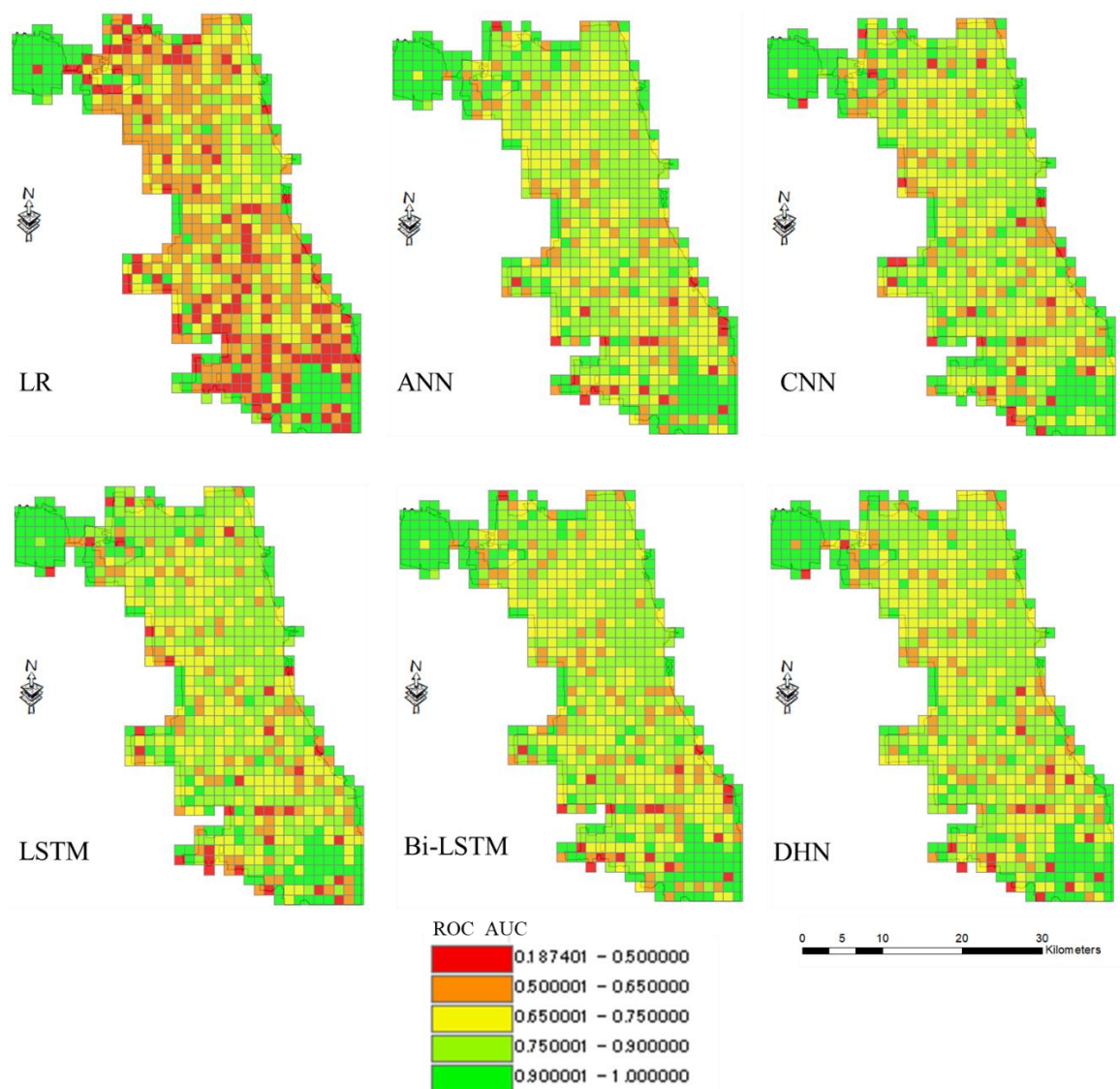
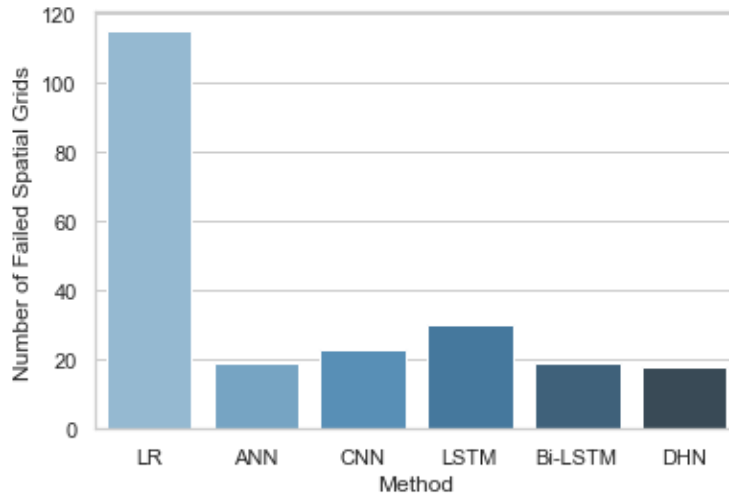


Figure 20: Spatial grid-level performance of methods in terms of AUC

The number of failed spatial grids for each method of analysis is indicated in (Figure 21).

There are a total of 710 spatial grids of 1kmx1km in Chicago city. LR has the highest

number of failed grids among these spatial grids, which is 115, while the DHN has the least number of failed grids (18).



**Figure 21: Number of failed spatial grids in spatiotemporal crash prediction for all methods**

According to Figure 20 and Figure 22, it can be seen that all of the classification methods have better prediction performance on central business district (CBD) than the rest of the city. Further, it also can be seen that the CBD has the highest crash frequency per grids and road density per grid. Thus, it can be concluded that spatial grids with the highest number of crashes and road density have better crash prediction compared to the rest of spatial grids.

Further, the spatial grids that did not have any crash in the study period also have good predictions. The spatial grids with no crash were predicted as no crash by all methods; thus, it resulted in perfect prediction.

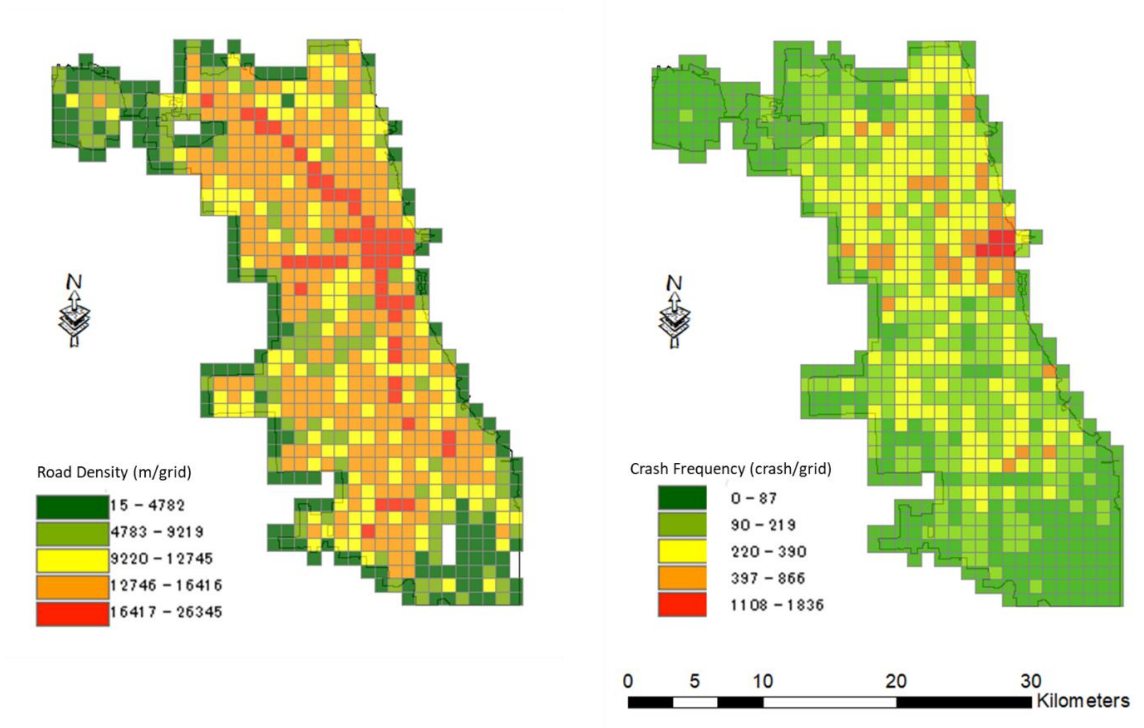


Figure 22: Road density and crash frequency for each spatial grid

To understand whether there is significant differences between the performance of every model, a significance test was conducted. In this study, the one-way analysis of variance (ANOVA) determines the significant differences between each model's performance. The models were tested ten times, and their AUC means were compared for the significance test.

The hypothesis test is written as follow:

$$\begin{cases} H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 \\ H_1: \text{There is a significant difference between means} \end{cases}$$

The significance test result indicates that the p-value of the test (p-value= 0.000) is less than the significance level ( $\alpha=0.05$ ); thus, it can be concluded that there is a significant difference between the mean AUC values of models.

Further, the pairwise performance of each pair of models was compared to determine if there is a significant difference between the performance of the models. For this purpose, the models' AUC values were compared using the Tukey test (Keselman & Rogan, 1977). Table 11 indicates the pairwise significance test of models. According to the table, there are insignificant differences in the AUC values of ANN, CNN, and LSTM with each other, while the rest of the pair of models' performance in terms of AUC are significantly different. Thus, it can be concluded that LR has the lowest performance in terms of AUC and the DHN has the highest performance in terms of AUC, while the ANN, CNN, LSTM have similar performance. Further, it can be concluded that Bi-LSTM has better performance in terms of AUC than LR, ANN, and LSTM and worse performance than the DHN.

**Table 11: Pairwise comparison of models performance using Tukey test for 1kmx1km spatial grids**

<b>Model</b>	<b>Compared with Model</b>	<b>p-value</b>	<b>Significance Difference Test</b>
LR	ANN	0.000	Significant
	CNN	0.000	Significant
	LSTM	0.000	Significant
	Bi-LSTM	0.000	Significant
	DHN	0.000	Significant
ANN	LR	0.000	Significant
	CNN	1.000	Insignificant
	LSTM	0.084	Insignificant
	Bi-LSTM	0.000	Significant
	DHN	0.000	Significant
CNN	LR	0.000	Significant
	ANN	1.000	Insignificant

	LSTM	0.041	Insignificant
	Bi-LSTM	0.000	Significant
	DHN	0.000	Significant
LSTM	LR	0.000	Significant
	ANN	0.084	Insignificant
	CNN	0.041	Insignificant
	Bi-LSTM	0.000	Significant
	DHN	0.000	Significant
Bi-LSTM	LR	0.000	Significant
	ANN	0.000	Significant
	CNN	0.000	Significant
	LSTM	0.000	Significant
	DHN	0.000	Significant
DHN	LR	0.000	Significant
	ANN	0.000	Significant
	CNN	0.000	Significant
	LSTM	0.000	Significant
	Bi-LSTM	0.000	Significant

#### 4.2.2 Results for 5kmX5km grids

In this section, the model predictions based on 5kmx5km spatial grids are discussed. Table 12 shows the models' performance based on the four performance measures of this study. The results indicate that for all of the four performance measures, the DHN has the best performance while LR has the worst. All other deep learning methods fall between these two methods.

**Table 12: Performance of six models comparison**

classifiers	AUC	Accuracy	Recall	False Alarm Rate	Training Time (s)
<b>LR</b>	0.8267	0.7441	0.7448	0.2561	<b>73</b>
<b>ANN</b>	0.8476	0.7673	0.7706	0.2336	136
<b>CNN</b>	0.8474	0.7679	0.7682	0.2322	189
<b>LSTM</b>	0.8477	0.7688	0.7618	0.2293	349

<b>Bi-LSTM</b>	0.8476	0.7686	0.7676	0.2311	375
<b>DHN</b>	<b>0.8586</b>	<b>0.7805</b>	<b>0.7679</b>	<b>0.2162</b>	242

Figure 23 indicates the AUC, recall, and accuracy for 5kmx5km spatial grid classifications. According to the figure, the AUC of the three deep learning methods (ANN, CNN, and LSTM) are similar. On the other hand, the LR has the lowest AUC. The DHN has outperformed all of the baseline methods in terms of AUC. Similarly, the accuracy of LS is the lowest among the six methods, while the DHN has the highest accuracy. In terms of recall, LR has the lowest performance while ANN, CNN, Bi-LSTM, and DHN have similar values. LSTM shows slightly worse performance in terms of recall compared to the rest of the deep learning methods.

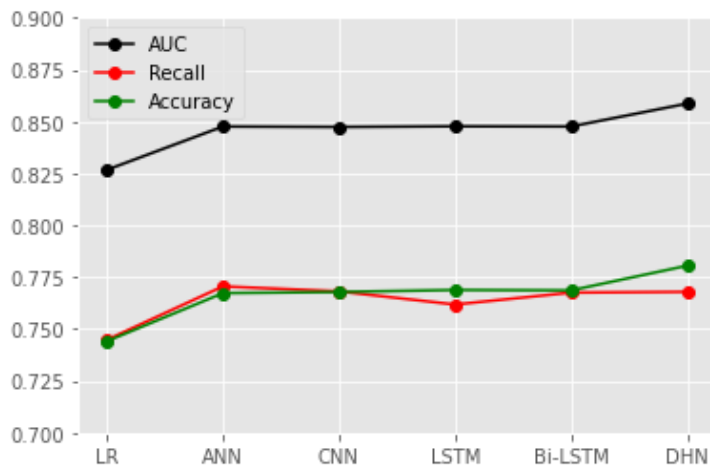


Figure 23: Comparison of AUC, recall, and accuracy for 5kmx5km spatial grids

Figure 24 shows the FAR of models for 5kmx5km spatial grids. The DHN outperformed the rest of the models in terms of FAR also. LR has the lowest performance in terms of

FAR, while the ANN, CNN, LSTM, and Bi-LSTM have similar FAR. It again confirms the superiority of the DHN method compared to the rest of the state-of-the-art methods.

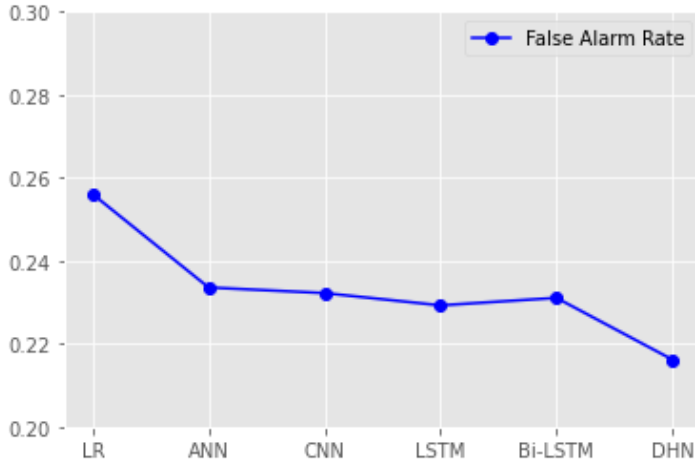


Figure 24: Comparison of FAR for 5kmx5km spatial grids

For evaluating models' consistency on the performance on different periods, the models' performance was tested the other two months of the year aside from the test set month (December). In the first case, the models were trained on the data from January to October and tested in November, and in the second case, the models were trained on the data from January to September and tested on October data. The models' performance results based on these two cases are indicated in Table 13. According to the table, the models' performance in the months of October and November is very close to the performance of models in December month. Therefore, it can be concluded that the models' performance is consistent each month.

Table 13: Models performance result for November and October as tests months

November				
classifiers	AUC	Accuracy	Recall	False Alarm Rate
LR	0.7681	0.6978	0.6967	0.3022
ANN	0.7880	0.7115	0.7153	0.2885

<b>CNN</b>	0.7866	0.7044	0.7181	0.2959
<b>LSTM</b>	0.7889	0.7153	0.7122	0.2847
<b>Bi-LSTM</b>	0.7884	0.7105	0.7146	0.2895
<b>DHN</b>	<b>0.7926</b>	<b>0.7165</b>	<b>0.7174</b>	<b>0.2835</b>
<b>October</b>				
<b>classifiers</b>	<b>AUC</b>	<b>Accuracy</b>	<b>Recall</b>	<b>False Alarm Rate</b>
<b>LR</b>	0.7724	0.6988	0.7010	0.3013
<b>ANN</b>	0.7888	0.7215	0.7015	0.2781
<b>CNN</b>	0.7875	0.7253	0.6966	0.2741
<b>LSTM</b>	0.7896	0.7196	0.7106	0.2803
<b>Bi-LSTM</b>	0.7888	0.7215	0.7015	0.2781
<b>DHN</b>	<b>0.7952</b>	<b>0.7345</b>	<b>0.6953</b>	<b>0.2647</b>

Figure 25 indicates the ROC curve of the models for 5kmx5km spatial grids. The ROC curve of LR captures the smallest area while it is the opposite for DHN. The three other deep learning models have nearly the same performance, and their ROC curves overlap accordingly.

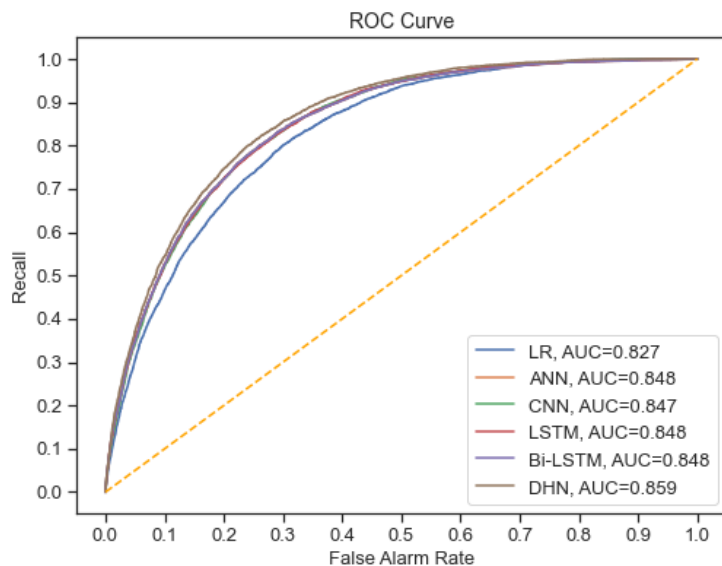
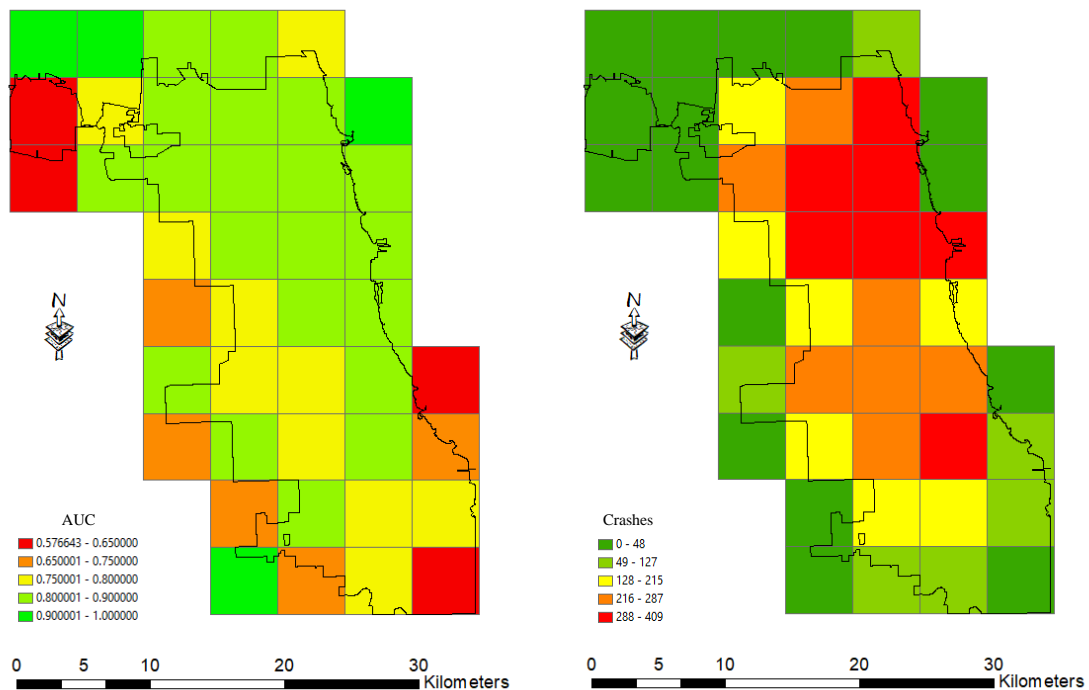


Figure 25: ROC curve for all methods with 5kmx5km spatial grids

The performance DHN model for each 5kmx5km spatial grid in terms of AUC is indicated in Figure 26. Overall, spatial grids' performance with a spatial resolution of 5kmx5km is better than 1kmx1km. The minimum AUC of 5kmx5km resolution spatial grid is about 0.58, which is better than random guess (0.5). Four spatial grids out of 43 have AUC between 0.58 and 0.65, and these spatial grids have the least number of crashes. Two spatial grids have AUCs between 0.9 and 1.0, and these two spatial grids also have the least number of crashes. Thus, the spatial grids with the least crashes have the least reliable prediction. The six spatial grids with the highest number of crashes perform well in terms of AUC, and they have AUC between 0.80 and 0.90. The rest of the spatial grids with a medium number of crashes have AUC between 0.65 to 0.90.



**Figure 26: Performance of each spatial grid (5kmx5km) and number of crashes in the test set (December)**

Similar to the 1kmx1km spatial resolution, the significance test was also conducted for 5kmx5km spatial resolution. The p-value of the test (p-value =0.0003) is less than the

significance level ( $\alpha=0.05$ ); thus, it can be concluded that there is a significant difference between the mean AUC values of models.

The pairwise mean AUC comparison test using the Tukey test is shown in Table 14. According to the table, there are insignificant differences in the AUC mean values of ANN, CNN, LSTM, and Bi-LSTM with each other, while the rest of the pair of models' performance in terms of AUC are significantly different.

**Table 14: Pairwise comparison of models performance using Tukey test for 5kmx5km spatial grids**

<b>Model</b>	<b>Compared with Model</b>	<b>p-value</b>	<b>Significance Difference Test</b>
LR	ANN	0.000	Significant
	CNN	0.000	Significant
	LSTM	0.000	Significant
	Bi-LSTM	0.000	Significant
	DHN	0.000	Significant
ANN	LR	0.000	Significant
	CNN	0.850	Insignificant
	LSTM	0.084	Insignificant
	Bi-LSTM	0.370	Insignificant
	DHN	0.000	Significant
CNN	LR	0.000	Significant
	ANN	0.850	Insignificant
	LSTM	0.041	Insignificant
	Bi-LSTM	0.090	Insignificant
	DHN	0.000	Significant
LSTM	LR	0.000	Significant
	ANN	0.084	Insignificant
	CNN	0.041	Insignificant
	Bi-LSTM	0.420	Insignificant
	DHN	0.000	Significant
Bi-LSTM	LR	0.000	Significant
	ANN	0.370	Insignificant
	CNN	0.090	Insignificant
	LSTM	0.420	Insignificant
	DHN	0.000	Significant
DHN	LR	0.000	Significant

	ANN	0.000	Significant
	CNN	0.000	Significant
	LSTM	0.000	Significant
	Bi-LSTM	0.000	Significant

### 4.3 Sensitivity analysis and feature importance

For understanding feature importance, a sensitivity analysis was conducted. In this analysis, the features were removed from the analysis one by one, and the effect of feature removal on model AUC was studied. This method was also used by (Moosavi et al., 2019).

The scenarios considered for sensitivity analysis as follows.

1. None: It means all features were included in the model.
2. Time: It means that the time feature (hour of day, day of week, and month of year) was removed from the model.
3. Holiday: The holiday data was removed from the model.
4. Weather: The weather condition data was removed from the model.
5. Taxi: The taxi trip data was removed from the model.
6. Bus: The bus ridership data category was removed from the model.
7. Red Light: Red light violations data
8. Speed: Speed violations data
9. Infrastructure: The infrastructure-related data was removed from the model (Note: this feature was only considered in DHN as an embedding layer as other models do not have such an embedding layer to consider the infrastructure inputs.).

### **4.3.1 Sensitivity analysis for spatial grids 1kmx1km**

The result of the sensitivity analysis for the six models is shown in Figure 27. According to the sensitivity analysis, the features' importance differs in models. According to Figure 27, time-related data is the most crucial variable in all models. Removal of time-related data results between 0.005 to 0.03 drop in the models. Holiday data seems to contribute to LSTM and Bi-LSTM models' performance; however, its significance on the rest of the models seems negligible.

On the other hand, weather data improves LR, ANN, and DHN model performance, while it does not affect the performance of CNN, LSTM, and Bi-LSTM models. Taxi data contribute to LR and ANN model performance, but its performance insignificant for the rest of the models. On the other hand, bus ridership, red light violations, and speed violations data seem to be insignificant. The infrastructure data importance for DHN looks pretty slight.

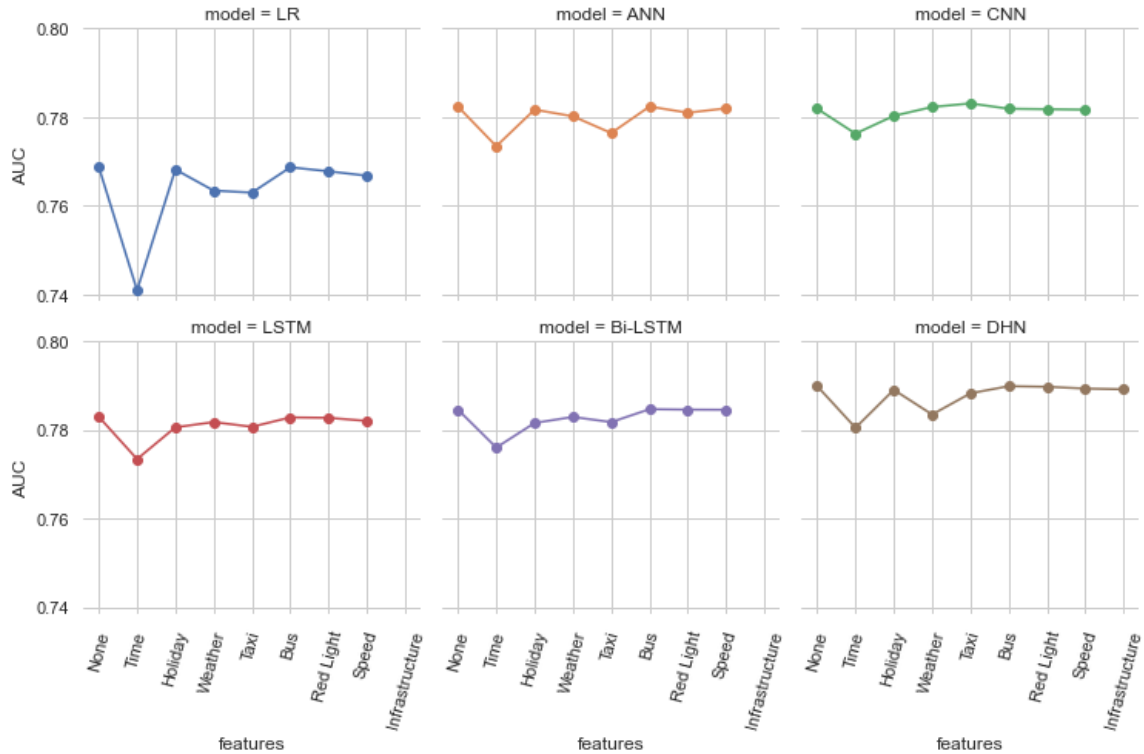


Figure 27: Sensitivity analysis for 1kmx1km spatial grids

### 4.3.2 Sensitivity analysis for spatial grids 5kmx5km

Sensitivity analysis was also considered for models with spatial grids 5kmx5km. Figure 28 shows the sensitivity analysis for 5kmx5km spatial grids. Similar to 1kmx1km spatial grids, the models' prediction performance is affected by different levels by variables inclusion. According to the figure, the time-related variables play a significant role in crash prediction for all models. Other features aside from the time of the crash are less critical than time variables. Weather conditions, taxi trips, and holiday data contribute to different levels of each model's performance. On the other hand, bus ridership, red light violations, and speed violations data seem insignificant in model prediction.

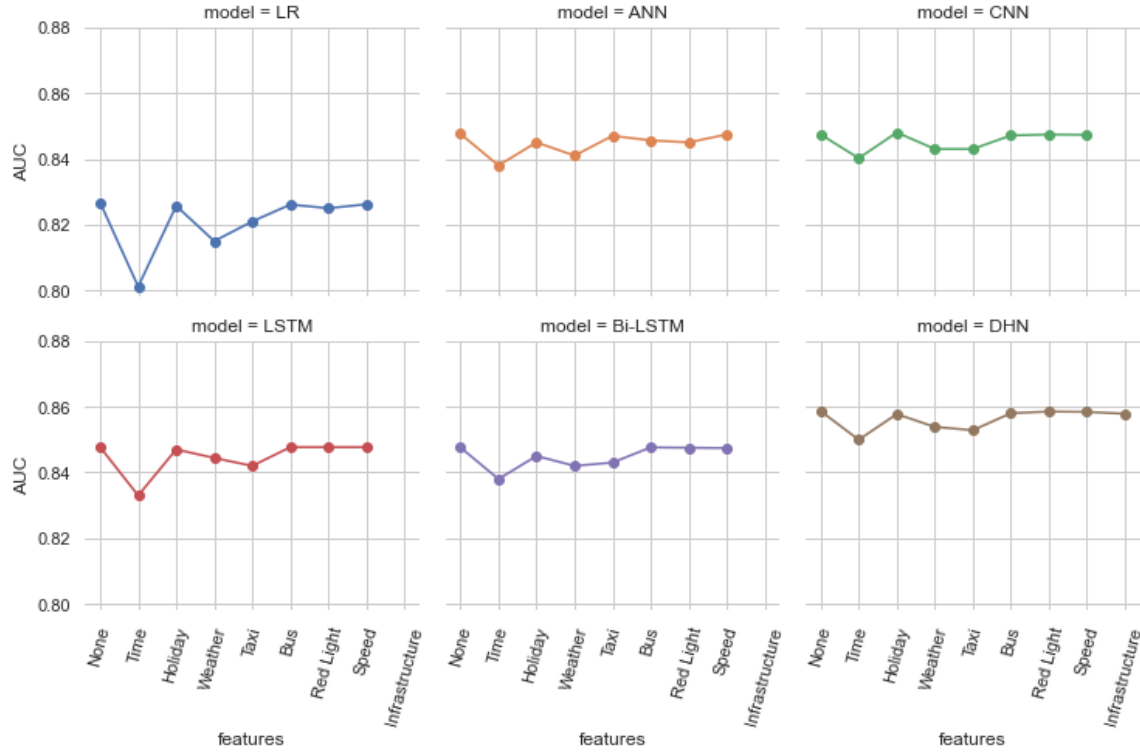


Figure 28: Sensitivity analysis for 5kmx5km spatial grids

#### 4.4 Comparison of different spatial resolution

Two spatial resolutions were considered for spatiotemporal crash prediction, which are 1kmx1km spatial grids and 5kmx5km spatial grids. In this subsection, the performance of the five baseline models and the proposed DHN method on these two spatial resolutions are compared. For comparison of spatial resolution of models, the AUC values of the models on the test set were implemented as it is the most stable performance among the performance measures. Figure 29 depicts the models' performance for the two spatial resolutions in terms of AUC. According to the figure, the models' performance differs on these two spatial resolutions, and the increase in performance for lower spatial resolution (5kmx5km spatial grids) in terms of AUC is consistent for all methods.

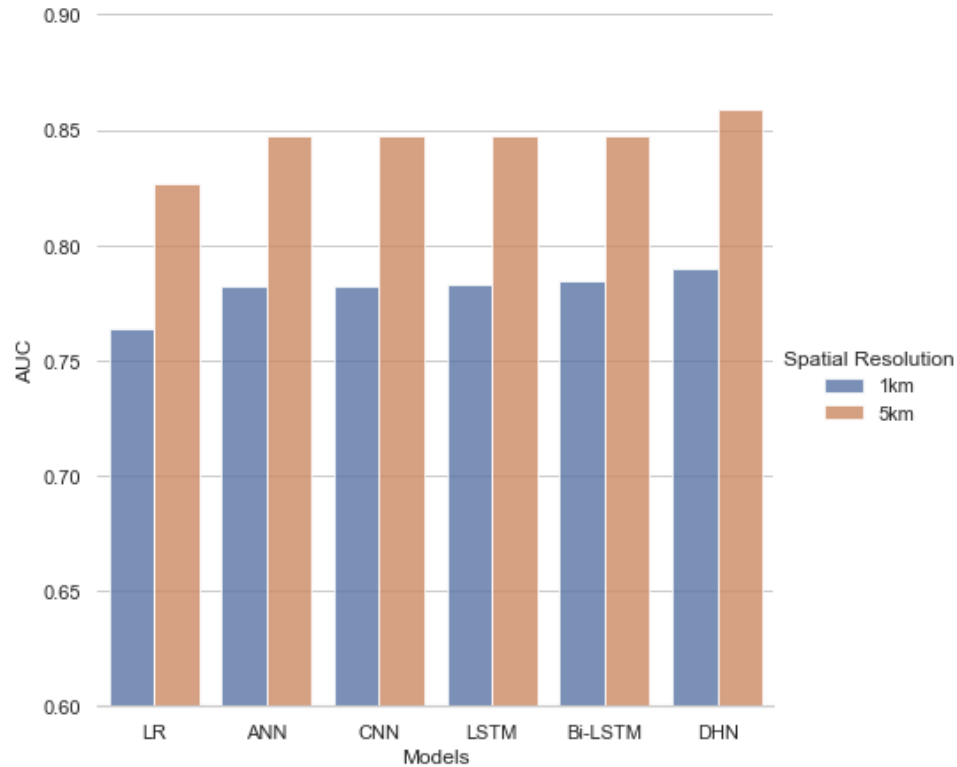


Figure 29: Comparison of models in different spatial resolutions in terms of AUC

As shown in Figure 26, the LR has the lowest performance for fine and granular spatial resolution, and the DHN has the highest performance in both spatial resolutions. Therefore, it can be concluded that the resolution of spatial grids has an effect on the prediction performance, and increasing the granularity of the spatial resolution leads to an increase in the models' performance accordingly.

#### 4.5 Discussion summary

The study results can be summarized as follows:

1. The spatial resolution of grids affects the model's performance. The granular spatial resolutions lead to higher prediction performance. This can be attributed to the rareness and randomness of a crash happening. As much as the spatial grids become

finer the number of crashes in each grid becomes less. This leads to an imbalance between crash and non-crash events that affects model prediction. Thus, the crash prediction performance decreases accordingly.

2. The proposed DHN has superior performance compared to the statistical LR and other deep learning methods. The DHN combines the ANN, CNN, and LSTM that capture spatial variability and temporal variability. The deep learning models produce similar prediction quality, while the statistical LR model has lower prediction quality than deep learning methods.
3. The time-related features have the highest significance on model performance. This can be attributed to human mobility changes over time. Human mobility significantly varies over time due to the necessity for travel each hour of the day. The level of human mobility has a direct relationship with the number of crash occurrences. Thus, the time-related facts affect the model prediction significantly. Other factors that affect the prediction of models are weather conditions, holidays data, infrastructure data, and these feature categories have different levels of impacts on each model. On the other hand, the red light violations, speed violations, and bus ridership are insignificant to the model prediction.
4. The taxi trips data and bus ridership data were thought to represent human mobility. However, these two features do not contribute to the models' prediction as expected. Human mobility leads to travel that directly influences road crash occurrences; however, the significance of taxi trips data and bus ridership data hint at such characteristics.

## CHAPTER 5

### CONCLUSION

#### 5.1 Conclusions

In this study, the short-term spatiotemporal crash prediction using deep learning methods was discussed. Various statistical and state-of-art deep learning methods were applied for crash prediction to compare the performance of each model in this case. Additionally, a deep learning method that combines different types of deep learning methods for crash prediction was applied. The result of the proposed DHN was compared to the statistical and deep learning method, and it was found that DHN outperforms the statistical and traditional deep learning models.

Two spatial resolutions for crash prediction were considered in this study, the spatial resolution of 1kmx1km and 5kmx5km. The models' prediction for these two spatial resolutions was studied separately. The model performance with the granular spatial resolution (5kmx5km) was better than the finer spatial resolution (1kmx1km) for all of the models. It seems that as much as the spatial resolution gets finer, the models' performance decreases. Additionally, the prediction performance for each spatial grid was evaluated. The spatial grids with the highest and lowest performance were highlighted.

Further, a sensitivity analysis was conducted to understand the importance of each variable in the models. The result of sensitivity analysis was compared for all methods, and it was found that crash time is the most crucial variable for crash prediction. Other variables such as holiday data, weather data, taxi trips data, and infrastructure data contributions to the

models differ from model to model. On the other hand, red-light violations, speed violations, and bus ridership data are insignificant to the model prediction.

## **5.2 Future Research Direction**

Various sources of data were used in this study to predict spatial grid-based crashes in Chicago city. Some data sources such as time of the crash, weather data, and taxi trips data. In the future, research is proposed to use some additional sources of big data to represent human mobility since human mobility is the predecessor of a road traffic crash. Some big data that may represent human mobility are cell phone data and social media data. These two sources of big data can be obtained from road users in real-time and incorporated in the crash prediction models accordingly. The inclusion of these data sources may help in higher model performance.

Further, the availability of road sensors data such as real-time traffic flow, occupancy rate, and speed in the short-term period (10 min, 15 min) for the entire city road network may help improve the prediction capability of the models. These data types are the actual representation of traffic state that would affect the occurrence of the crash; thus, they would have a significant impact on the models' performance. Additionally, crash detection by road sensors in freeways and expressways has already attracted the attention of many researchers. The crash detection by the sensor is also applied to intersection and weaving segments by some authors. Therefore, it would be interesting to consider the sensor data combined with the big data such as cell phone data in addition to the data used in this research for crash prediction.

## References

- Abdel-Aty, M. A., Hassan, H. M., Ahmed, M., & Al-Ghamdi, A. S. (2012). Real-time prediction of visibility related crashes. *Transportation Research Part C: Emerging Technologies*, 24, 288–298. <https://doi.org/10.1016/j.trc.2012.04.001>
- Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, M. F., & Hsia, L. (2004). Predicting freeway crashes from loop detector data by matched case-control logistic regression. *Transportation Research Record*, 1897, 88–95. <https://doi.org/10.3141/1897-12>
- Aguero-Valverde, J., & Jovanis, P. P. (2006). Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis and Prevention*, 38(3), 618–625. <https://doi.org/10.1016/j.aap.2005.12.006>
- Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.
- Amoros, E., Martin, J. L., & Laumon, B. (2003). Comparison of road crashes incidence and severity between some French counties. *Accident Analysis and Prevention*, 35(4), 537–547. [https://doi.org/10.1016/S0001-4575\(02\)00031-3](https://doi.org/10.1016/S0001-4575(02)00031-3)
- Arguez, A., Durre, I., Applequist, S., Squires, M., Vose, R., Yin, X., & Bilotta, R. (2020). *NOAA's U.S. Climate Normals (1981-2010). Hourly Weather*. NOAA National Centers for Environmental Information. <https://www.ncdc.noaa.gov/cdo-web/datatools/lcd>

- Bao, J., Liu, P., Qin, X., & Zhou, H. (2018). Understanding the effects of trip patterns on spatially aggregated crashes with large-scale taxi GPS data. *Accident Analysis and Prevention*, *120*(September), 281–294. <https://doi.org/10.1016/j.aap.2018.08.014>
- Bao, J., Liu, P., & Ukkusuri, S. V. (2019). A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data. *Accident Analysis and Prevention*, *122*(November 2018), 239–254. <https://doi.org/10.1016/j.aap.2018.10.015>
- Bao, J., Liu, P., Yu, H., & Xu, C. (2017). Incorporating twitter-based human activity information in spatial analysis of crashes in urban areas. *Accident Analysis and Prevention*, *106*(July), 358–369. <https://doi.org/10.1016/j.aap.2017.06.012>
- Basso, F., Basso, L. J., Bravo, F., & Pezoa, R. (2018). Real-time crash prediction in an urban expressway using disaggregated data. *Transportation Research Part C: Emerging Technologies*. <https://doi.org/10.1016/j.trc.2017.11.014>
- Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning Long-Term Dependencies with Gradient Descent is Difficult. *IEEE Transactions on Neural Networks*. <https://doi.org/10.1109/72.279181>
- Chang, D. C., Eastman, B., Talamini, M. A., Osen, H. B., Tran Cao, H. S., & Coimbra, R. (2011). Density of surgeons is significantly associated with reduced risk of deaths from motor vehicle crashes in US counties. *Journal of the American College of Surgeons*, *212*(5), 862–866. <https://doi.org/10.1016/j.jamcollsurg.2011.01.057>

- Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171–209. <https://doi.org/10.1007/s11036-013-0489-0>
- Chen, Q., Song, X., Yamada, H., & Shibasaki, R. (2016). Learning deep representation from big and heterogeneous data for traffic accident inference. *30th AAAI Conference on Artificial Intelligence, AAAI 2016*, 338–344.
- Chicago Data Portal. (2020). *Traffic Crashes - Crashes | City of Chicago | Data Portal*. <https://data.cityofchicago.org/Transportation/Traffic-Crashes-Crashes/85ca-t3if>
- Continental Motors. (2020). *How many cars are in Chicago? Commuter Vehicle Numbers*. <https://www.continentalmotors.com/blog/how-many-cars-are-in-chicago/>
- Darwiche, A. L. (2009). a Gis Safety Study and a County-Level Spatial Analysis of Crashes in the State of Florida. *University of Central Florida*, 2(2009), 2004–2019.
- Data Commons. (2020). *Chicago - Place Explorer - Data Commons*. <https://datacommons.org/place/geoId/1714000>
- Donaldson, A. E., Cook, L. J., Hutchings, C. B., & Dean, J. M. (2006). Crossing county lines: The impact of crash location and driver’s residence on motor vehicle crash fatality. *Accident Analysis and Prevention*, 38(4), 723–727. <https://doi.org/10.1016/j.aap.2006.01.002>
- Fischer, J. W., Resources Science, & Division, I. (2005). *Safe, Accountable, Flexible, Efficient Transportation Equity Act--A Legacy for Users (SAFETEA-LU Or SAFETEA): Selected Major Provisions*.

- Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, *12*(10), 2451–2471.  
<https://doi.org/10.1162/089976600300015015>
- Hanna, C. L., Laflamme, L., & Bingham, C. R. (2012). Fatal crash involvement of unlicensed young drivers: County level differences according to material deprivation and urbanicity in the United States. *Accident Analysis and Prevention*, *45*, 291–295.  
<https://doi.org/10.1016/j.aap.2011.07.014>
- Hochreiter, S., & Schmidhuber, J. (1997a). Long Short-Term Memory. *Neural Computation*. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hochreiter, S., & Schmidhuber, J. (1997b). Long Short-Term Memory. *Neural Computation*. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hochreiter, S., & Schmidhuber, J. (1997c). Long Short-Term Memory. *Neural Computation*. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Huang, H., Abdel-Aty, M. A., & Darwiche, A. L. (2010a). County-level crash risk analysis in Florida: Bayesian spatial modeling. *Transportation Research Record*, *2148*, 27–37.  
<https://doi.org/10.3141/2148-04>
- Huang, H., Abdel-Aty, M. A., & Darwiche, A. L. (2010b). County-level crash risk analysis in Florida: Bayesian spatial modeling. *Transportation Research Record*, *2148*, 27–37.  
<https://doi.org/10.3141/2148-04>
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260.

- Karlaftis, M. G., & Tarko, A. P. (1998). Heterogeneity considerations in accident modeling. *Accident Analysis and Prevention*, 30(4), 425–433.  
[https://doi.org/10.1016/S0001-4575\(97\)00122-X](https://doi.org/10.1016/S0001-4575(97)00122-X)
- Keselman, H. J., & Rogan, J. C. (1977). The Tukey multiple comparison test: 1953--1976. *Psychological Bulletin*, 84(5), 1050.
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.  
<https://doi.org/10.1038/nature14539>
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*. <https://doi.org/10.1162/neco.1989.1.4.541>
- Li, P., Abdel-Aty, M., & Yuan, J. (2020). Real-time crash risk prediction on arterials based on LSTM-CNN. *Accident Analysis and Prevention*, 135(July 2019), 105371.  
<https://doi.org/10.1016/j.aap.2019.105371>
- Liu, G., & Guo, J. (2019). Bidirectional LSTM with attention mechanism and convolutional layer for text classification. *Neurocomputing*.  
<https://doi.org/10.1016/j.neucom.2019.01.078>
- Maciag, M. (2017). *Vehicle Ownership in U.S. Cities Data and Map*.  
<https://www.governing.com/gov-data/car-ownership-numbers-of-vehicles-by-city-map.html>

- Mannering, F. L., Shankar, V., & Bhat, C. R. (2016). Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research, 11*, 1–16. <https://doi.org/10.1016/j.amar.2016.04.001>
- Marsland, S. (2015). *Machine learning: an algorithmic perspective*. CRC press.
- Martin, P. T. (2003). *Detector Technology Evaluation (MPC-03-154)* (Issue November).
- Michie, D., Spiegelhalter, D. J., Taylor, C. C., & others. (1994). Machine learning. *Neural and Statistical Classification, 13*(1994), 1–298.
- Moosavi, S., Samavatian, M. H., Parthasarathy, S., Teodorescu, R., & Ramnath, R. (2019). Accident risk prediction based on heterogeneous sparse data: New dataset and insights. *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems, 33–42*. <https://doi.org/10.1145/3347146.3359078>
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.
- Noland, R. B. (2003). Traffic fatalities and injuries: The effect of changes in infrastructure and other trends. *Accident Analysis and Prevention, 35*(4), 599–611. [https://doi.org/10.1016/S0001-4575\(02\)00040-4](https://doi.org/10.1016/S0001-4575(02)00040-4)
- Noland, R. B., & Oh, L. (2004). The effect of infrastructure and demographic change on traffic-related fatalities and crashes: A case study of Illinois county-level data. *Accident Analysis and Prevention, 36*(4), 525–532. [https://doi.org/10.1016/S0001-4575\(03\)00058-7](https://doi.org/10.1016/S0001-4575(03)00058-7)
- Núñez, J. C., Cabido, R., Pantrigo, J. J., Montemayor, A. S., & Vélez, J. F. (2018). Convolutional Neural Networks and Long Short-Term Memory for skeleton-based

- human activity and hand gesture recognition. *Pattern Recognition*.  
<https://doi.org/10.1016/j.patcog.2017.10.033>
- Olah, C. (2015). *Understanding LSTM Networks -- colah's blog*.  
<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Ozenne, B., Subtil, F., & Maucort-Boulch, D. (2015). The precision-recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *Journal of Clinical Epidemiology*, 68(8), 855–859.  
<https://doi.org/10.1016/j.jclinepi.2015.02.010>
- Park, S. hun, Kim, S. min, & Ha, Y. guk. (2016). Highway traffic accident prediction using VDS big data analysis. *Journal of Supercomputing*, 72(7), 2815–2831.  
<https://doi.org/10.1007/s11227-016-1624-z>
- Parsa, A. B., Taghipour, H., Derrible, S., & Mohammadian, A. (Kouros). (2019). Real-time accident detection: Coping with imbalanced data. *Accident Analysis and Prevention*, 129(January), 202–210. <https://doi.org/10.1016/j.aap.2019.05.014>
- Ren, H., Song, Y., Wang, J., Hu, Y., & Lei, J. (2018). A Deep Learning Approach to the Citywide Traffic Accident Risk Prediction. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 2018-Novem*, 3346–3351.  
<https://doi.org/10.1109/ITSC.2018.8569437>
- Rhee, K. A., Kim, J. K., Lee, Y. I., & Ulfarsson, G. F. (2016). Spatial regression analysis of traffic crashes in Seoul. *Accident Analysis and Prevention*, 91, 190–199.  
<https://doi.org/10.1016/j.aap.2016.02.023>

- Schaffer, S. (2019). *Chicago to Tackle Congestion and Transportation Inequity* | NRDC.  
<https://www.nrdc.org/experts/chicago-tackle-congestion-and-transportation-inequity>
- Shi, Q., & Abdel-Aty, M. (2015). Big Data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. *Transportation Research Part C: Emerging Technologies*, 58, 380–394.  
<https://doi.org/10.1016/j.trc.2015.02.022>
- Stacy, C. D., & Robert, G. B. (2020). *Data - Transportation Energy Data Book*  
*Transportation Energy Data Book*. <https://tedb.ornl.gov/data/>
- Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*. [https://doi.org/10.1016/S0034-4257\(97\)00083-7](https://doi.org/10.1016/S0034-4257(97)00083-7)
- Tian, C., Ma, J., Zhang, C., & Zhan, P. (2018). A deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network. *Energies*, 11(12). <https://doi.org/10.3390/en11123493>
- Train, K. E. (2003). Discrete choice methods with simulation. In *Discrete Choice Methods with Simulation*. <https://doi.org/10.1017/CBO9780511753930>
- Traynor, T. L. (2008). Regional economic conditions and crash fatality rates - a cross-county analysis. *Journal of Safety Research*, 39(1), 33–39.  
<https://doi.org/10.1016/j.jsr.2007.10.008>

- United Nations. (2014). World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352). In *New York, United.*  
<https://doi.org/10.4054/DemRes.2005.12.9>
- Wang, L., Abdel-Aty, M., Shi, Q., & Park, J. (2015). Real-time crash prediction for expressway weaving segments. *Transportation Research Part C: Emerging Technologies*, *61*, 1–10. <https://doi.org/10.1016/j.trc.2015.10.008>
- Washington, S., van Schalkwyk, I., You, D., Shin, K., & Samuelson, J. P. (2010). *Plansafe: forecasting the safety impacts of socio-demographic changes and safety countermeasures.*
- WHO. (2004). *World report on road traffic injury prevention.*  
<https://www.who.int/publications/i/item/world-report-on-road-traffic-injury-prevention>
- WHO. (2018). WHO | Global status report on road safety 2018. In WHO. World Health Organization.
- WISQARS. (2020). *WISQARS (Web-based Injury Statistics Query and Reporting System)/Injury Center/CDC.* <https://www.cdc.gov/injury/wisqars/>
- Xie, K., Ozbay, K., Kurkcu, A., & Yang, H. (2017). Analysis of Traffic Crashes Involving Pedestrians Using Big Data: Investigation of Contributing Factors and Identification of Hotspots. *Risk Analysis*, *37*(8), 1459–1476. <https://doi.org/10.1111/risa.12785>

- Xu, C., Liu, P., Wang, W., & Li, Z. (2012). Evaluation of the impacts of traffic states on crash risks on freeways. *Accident Analysis and Prevention*, *47*, 162–171. <https://doi.org/10.1016/j.aap.2012.01.020>
- Xu, D., Shi, Y., Tsang, I. W., Ong, Y. S., Gong, C., & Shen, X. (2020). Survey on Multi-Output Learning. *IEEE Transactions on Neural Networks and Learning Systems*, *31*(7), 2409–2429. <https://doi.org/10.1109/TNNLS.2019.2945133>
- Yan, X., & Wu, J. (2014). Effectiveness of variable message signs on driving behavior based on a driving simulation experiment. *Discrete Dynamics in Nature and Society*, *2014*. <https://doi.org/10.1155/2014/206805>
- Yu, R., & Abdel-Aty, M. (2013). Utilizing support vector machine in real-time crash risk evaluation. *Accident Analysis and Prevention*. <https://doi.org/10.1016/j.aap.2012.11.027>
- Yuan, J., Abdel-Aty, M., Gong, Y., & Cai, Q. (2019). Real-Time Crash Risk Prediction using Long Short-Term Memory Recurrent Neural Network. *Transportation Research Record*, *2673*(4), 314–326. <https://doi.org/10.1177/0361198119840611>
- Zhao, Z., Chen, W., Wu, X., Chen, P. C. V., & Liu, J. (2017). LSTM network: A deep learning approach for short-term traffic forecast. *IET Image Processing*. <https://doi.org/10.1049/iet-its.2016.0208>
- Zheng, X., & Chen, W. (2021). An Attention-based Bi-LSTM Method for Visual Object Classification via EEG. *Biomedical Signal Processing and Control*, *63*(August 2020), 102174. <https://doi.org/10.1016/j.bspc.2020.102174>

- Zheng, Zibin, Yang, Y., Liu, J., Dai, H. N., & Zhang, Y. (2019). Deep and Embedded Learning Approach for Traffic Flow Prediction in Urban Informatics. *IEEE Transactions on Intelligent Transportation Systems*.  
<https://doi.org/10.1109/TITS.2019.2909904>
- Zheng, Zuduo, Ahn, S., & Monsere, C. M. (2010). Impact of traffic oscillations on freeway crash occurrences. *Accident Analysis and Prevention*, 42(2), 626–636.  
<https://doi.org/10.1016/j.aap.2009.10.009>
- Zhu, L., Li, T., & Du, S. (2019). TA-STAN: A Deep Spatial-Temporal Attention Learning Framework for Regional Traffic Accident Risk Prediction. *Proceedings of the International Joint Conference on Neural Networks, 2019-July(July)*, 1–8.  
<https://doi.org/10.1109/IJCNN.2019.8852212>

## Vitae

Mohammad Tamim Kashifi has completed his BSc in civil engineering at Kabul University, Afghanistan, in 2014. Additionally, he has completed his MSc in civil and environmental engineering at King Fahd University of Petroleum & Mineral (KFUPM), Saudi Arabia, in 2021. He worked as a traffic and transportation engineer in various infrastructure projects in Afghanistan between 2014 and 2019. His research interests are traffic engineering, machine learning and deep learning applications in traffic engineering, and traffic safety engineering. He is currently working on spatiotemporal traffic crash modeling and prediction using hybrid deep learning methods, statistical traffic crash analysis, crash severity prediction, traffic state prediction for Intelligent Transportation System (ITS) applications, and similar projects. Further, he has two journal articles titled “Transit leverage assessment and climate change mitigation pathway for urbanized areas” published in the International Journal of Global Warming and “A review of factors and benefits of non-motorized transport: a way forward for developing countries” published in Environment, Development and Sustainability journal.

Name	:{Mohammad Tamim Kashifi }
Nationality	:{Afghanistan }
Date of Birth	:{8/31/1992}
Email	:{tamim_motebahher@yahoo.com}
Cell Phone	:{+93744816693 }

Address : [P. O, Baraki Square, Kabul 1003, Afghanistan]

Academic Background : [Traffic Engineering]