DECLARATION

The work provided in this thesis, unless otherwise referenced, is the researcher's own work, and has not been submitted elsewhere for any other degree or qualification.

Student's name: Ibrahim ElAstal
Signature: [signature]
Date: 21/01/2015
Modeling Road Accident Black Spots in Gaza Strip Using Artificial Neural Network (ANN)  
2000-2013

Ibrahim M. A. ElAstal

Supervised by
Yahya R. Sarraj
Associated Professor of Transportation

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master in Civil Engineering, Infrastructure Management.
نتيجة الحكم على أطروحة ماجستير

بناءً على موافقة شئون البحث العلمي والدراسات العليا بالجامعة الإسلامية بغزة على تشكيك لجنة الحكم على أطروحة الباحث/ إبراهيم محمد عبد السلام الأسطل لنقل درجة الماجستير في كلية الهندسة قسم الهندسة المدنية- اللبنى التحتية وموضوعها:

نمذجة النقاط السوداء لحوادث الطرق في قطاع غزة باستخدام شبكات الأعصاب الإصطناعية 2000-2013

Modeling Road Accident Black Spots in Gaza Strip Using Artificial Neural Networks (ANN) 2000-2013

وبعد المناقشة التي تمت اليوم الأحد 8 صفر 1436هـ الموافق 30/11/2014م الساعة الثانية عشرة ظهراً، اجتمعت لجنة الحكم على الأطروحة والمكونة من:

د. حبيبي رشدي السراج

د. عصام حسين المصري

د. علاء الدين داوود الجماصي

د. عيسى ابراهيم المناقشة الداخلية

وأصدرت لجنة الحكم على أطروحة ماجستير قسم الهندسة المدنية، اللبنى التحتية، ود. محمد عبد السلام الأسطل درجة ماجستير في كلية الهندسة قسم الهندسة المدنية، اللبنى التحتية.

والله وإلى الفاتي...

مساءلة نائب الرئيس للبحث العلمي والدراسات العليا

أ. فواز علي العاجز
بسم الله الرحمن الرحيم

"قل إني صلاني ونيسكي ومحياي ومماتي لله رب العالمين"

صدق الله العظيم

الأنعام-آية 162
DEDICATION

TO MY FAMILY
I DEDICATE THIS WORK.
ACKNOWLEDGEMENT

First of all, all thanks and appreciations go to Allah for His unlimited blessings and for giving me the strength to complete this work.

Special Thanks are due to my supervisor Dr. Yahya Sarraj as he was patient and advisable to me.

Also, I like to give special thanks to the Ministry of Transportation, Gaza Police Department, Gaza Strip municipalities, my students at the Islamic University of Gaza, my colleagues, Mr. Shukri AlAstal and Mr. Mahmoud Shikh Eid.

Finally, my sincere thanks are due to all people who supported me to complete this work.
ABSTRACT

Around the world, traffic accidents are considered as one of the most important causes of death. In Palestine, recorded traffic accidents increased twice in the period from 2007 to 2013. This situation reflects the dangerous situation on the roads.

In this research, analysis of the traffic accidents in Palestine from 2000 to 2013 is carried out. It shows that accidents in Palestine have a non-stable behavior due to the continuous changing in the political situation.

The databases and Gaza Aerial photo are collected from Ministry of Transportation, Land Authority and Gaza Police Department.

A team consisting of 20 members, covering all Gaza strip governorates, worked to specify the location of the accidents from 2000 to 2005 using ArcGIS. They located 67% of the accidents. These data are used to identify black spots and to find the most dangerous locations in Gaza Strip.

The black spot identification shows that Salah Aldeen Street has more than five black spots. Therefore, it is recommended to apply the Rout Action Approach in order to prepare an accident reduction plan.

After that, an artificial neural network model is developed using spots factors and weighting. The main factors that are used in the model are the traffic volume, the number of the intersections in the spot, the existence of the median in the main road, the existence of one-way roads in the spot, the surface type, the design speed, the number of lanes and the road design width. In validation process, the model sensitivity value is (91%), model specificity value is (72%), and the coefficient of determination ($R^2$) value is 0.55. These numbers indicate that the model could represent the real data.

Keywords:
Black Spot, Hot Spot, Traffic accidents analysis, Artificial Neural Networks, ANN, Gaza, Palestine.
تعتبر الحوادث المرورية أحد أهم أسباب الوفاة حول العالم. في فلسطين، الحوادث المرورية المسجلة زادت في الفترة من 2007 وحتى 2013. هذا الوضع يعكس ضعف الطرق على الطرق.

يقوم البحث بتحليل عدد الحوادث المرورية في فلسطين، في الفترة من 2000 وحتى 2013. وقد تم التوصل إلى أن الحوادث المرورية في فلسطين غير مستقرة وذلك للعنصر 정치ي في الأوضاع السياسية.

تم الحصول على قواعد بيانات الحوادث المرورية والطرق والصور الجوية من قبل كل من شرطة غزة، وزارة النقل، وحركة الأراضي، وشرطة ممرور غزة.


وبعد عملية تعريف النقاط السوداء، تم توقع أكثر من خمس نقاط سوداء على طول شارع صلاح الدين، لوضع خطة للحد من الحوادث المرورية.

في عملية التحقق من صحة النموذج تم إيجاد أن معامل الحساسية النموذج 10% بينما معامل الخاصية 27% ومعامل التحديد $R^2$ 0.55. هذه القيم تعطي مؤشر على أن النموذج يمثل البيانات الواقعية.

كلمات مفتاحية:
النقاط السوداء، النقاط الساخنة، تحليل الحوادث المرورية، الشبكات الصناعية العصبية، غزة، فلسطين.
# LIST OF CONTENTS

DEDICATION .................................................................................................................. I

ACKNOWLEDGEMENT .................................................................................................. II

ABSTRACT ....................................................................................................................... III

ARABIC ABSTRACT .......................................................................................................... IV

LIST OF CONTENTS ........................................................................................................ V

LIST OF ABBREVIATIONS ............................................................................................... VII

LIST OF TABLES ............................................................................................................... VIII

LIST OF FIGURES ............................................................................................................ IX

1 INTRODUCTION ......................................................................................................... 1

1.1 Problem Statement .................................................................................................. 1

1.2 Research Objectives .............................................................................................. 1

1.3 Research Methodology .......................................................................................... 2

1.4 Research Structure ................................................................................................ 2

2 TRAFFIC ACCIDENTS ANALYSIS ............................................................................. 4

2.1 Traffic accident definitions ....................................................................................... 4

2.2 History of traffic accidents ....................................................................................... 4

2.3 Factors contributing to traffic accidents ................................................................... 5

2.4 Traffic accidents classifications ............................................................................... 6

2.5 Traffic accidents casualties ..................................................................................... 7

2.6 Traffic safety measurements .................................................................................... 9

3 DATA MINING ........................................................................................................... 11

3.1 DIKW hierarchy ..................................................................................................... 11

3.2 Definition of data mining ....................................................................................... 12

3.3 Knowledge Discovery from Data steps (KDD steps) ............................................... 12

3.4 Data mining functionalities ..................................................................................... 13

3.5 Mining spatial and spatiotemporal data .................................................................. 14

3.6 Data mining in traffic accidents analysis ................................................................. 15

4 BLACK SPOT ANALYSIS ........................................................................................... 16

4.1 Accident investigation and reduction ....................................................................... 16

4.2 Black spot identification, ranking and selection ...................................................... 17

4.3 Black spot modeling ............................................................................................... 19

4.4 Using Geographic Information System (GIS) in black spot analysis ................. 21

V
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AADT</td>
<td>Average Annual Daily Traffic</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Network</td>
</tr>
<tr>
<td>CAD</td>
<td>Computer Aided Design</td>
</tr>
<tr>
<td>CR</td>
<td>Causality Rate</td>
</tr>
<tr>
<td>CSO</td>
<td>Central Statistics Office</td>
</tr>
<tr>
<td>DALYs</td>
<td>Disability Adjusted Life Years</td>
</tr>
<tr>
<td>DBMS</td>
<td>Data Base Management System</td>
</tr>
<tr>
<td>DfT</td>
<td>Department For Transportation</td>
</tr>
<tr>
<td>DIKW Hierarchy</td>
<td>Data-Information-Knowledge-Wisdom Hierarchy</td>
</tr>
<tr>
<td>DPD</td>
<td>Denver Police Department</td>
</tr>
<tr>
<td>GA</td>
<td>Genetic Algorithm</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross Domestic Product</td>
</tr>
<tr>
<td>GIS</td>
<td>Geographic Information System</td>
</tr>
<tr>
<td>GPD</td>
<td>Gaza Police Department</td>
</tr>
<tr>
<td>HIG</td>
<td>Hanover Insurance Group</td>
</tr>
<tr>
<td>IRTAD</td>
<td>International Road Traffic and Accident Database</td>
</tr>
<tr>
<td>KDD</td>
<td>Knowledge Discover from Data</td>
</tr>
<tr>
<td>KSA</td>
<td>Kingdom of Saudi Arabia</td>
</tr>
<tr>
<td>MOT</td>
<td>Ministry Of Transportation</td>
</tr>
<tr>
<td>MPH</td>
<td>Mile Per Hour</td>
</tr>
<tr>
<td>PCBS</td>
<td>Palestinian Central Bureau of Statistics</td>
</tr>
<tr>
<td>PDO</td>
<td>Property Damage Only</td>
</tr>
<tr>
<td>PIA</td>
<td>Personal Injury Accident</td>
</tr>
<tr>
<td>RDBMS</td>
<td>Relational Data Base Management System</td>
</tr>
<tr>
<td>TxDOT</td>
<td>Texas Department Of Transportation</td>
</tr>
<tr>
<td>UAE</td>
<td>United Arab Emirate</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>UN</td>
<td>United Nation</td>
</tr>
<tr>
<td>UNECE</td>
<td>United Nations Economic Commission for Europe</td>
</tr>
<tr>
<td>USA</td>
<td>United State of America</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table (2.1): Leading causes of death, 2004 versus 2030 (WHO, 2012)------------------ 5
Table (2.2): Different exposures used in the safety measurements (Shinar 2007). - 10
Table (6.1): Traffic Accidents and Casualties in Palestine. ------------------------ 28
Table (6.2): Traffic Accidents and Casualties in Gaza. -------------------------- 28
Table (6.3): Accidents Severity in Gaza Strip (2000-2005)---------------------- 33
Table (6.4): Accidents during the day (2000-2005) ---------------------------- 33
Table (6.5): Accidents involved carts, motorcycles and bicycles (2000-2005) ----- 33
Table (6.6): Vehicles classification due to Insurance status (2000-2005)-------- 33
Table (6.7): Accidents According to Injuries Age------------------------------- 34
Table (7.1): Accidents in the Database---------------------------------------- 37
Table (7.2): Accidents in the Database---------------------------------------- 37
Table (7.3): The most dangerous black spot in Gaza Strip from 2000 to 2005----- 39
Table (7.4): Number of intersection factor ------------------------------------ 39
Table (7.5): Flow factor ------------------------------------------------------ 40
Table (7.6): Median factor ---------------------------------------------------- 40
Table (7.7): Existence of one-way road factor ------------------------------- 40
Table (7.8): Surface factor -------------------------------------------------- 40
Table (7.9): Running model details ------------------------------------------ 42
LIST OF FIGURES

Figure (1.1): Distribution of global injury mortality by cause (WHO, 2012),........ 1
Figure (2.1): Factors contributing to traffic accidents...................................... 6
Figure (3.1): DIKW hierarchy (Ackoff, 2009).................................................. 11
Figure (5.1): Typical Structure for ANN (An-Najjar, 2005)............................... 24
Figure (5.2): The Concept of ANN (An-Najjar, 2005)...................................... 25
Figure (6.1): Adjudge accidents vs amicably resolved accidents in Gaza.......... 29
Figure (6.2): Causality Rate (CR) in Palestine.................................................. 30
Figure (6.3): Causality Rate (CR) in Gaza....................................................... 30
Figure (6.4): Fatal Causality Rate (FCR) in Palestine ..................................... 30
Figure (6.5): Fatal Causality Rate (FCR) in Gaza............................................ 31
Figure (6.6): Fatality Risk per 100,000 Inhabitants in 2011.............................. 31
Figure (6.7): Fatality Risk per 10,000 Vehicles in 2011.................................... 31
Figure (6.8): Accidents in Gaza Strip.............................................................. 32
Figure (6.9): The Summation of Monthly Accident in Gaza Strip (2000-2005) .... 34
Figure (7.1): ANN in VGD Program .............................................................. 41
Figure (7.2): Topology of the Model .................................................................. 41
Figure (7.3): Network Map Analysis ............................................................... 42
Figure (7.4): Neural network Black Spot weight .............................................. 43
1 INTRODUCTION

This chapter discusses problem statement, objectives, brief methodology and research structure.

1.1 Problem Statement

Traffic accidents are considered one of the main issues facing societies around the world where they consume human and financial resources. World Health Organization (WHO) statistics show that each year traffic accidents kill more than 1.2 million and injure (20 to 50) million. Moreover, they cause losses with more than 518 Billion American dollar around the world, this number represent from 1% to 3% of the Gross Domestic Product (GDP) (WHO, 2012).

Figure (1.1) shows that traffic accidents caused about 22% of the injury mortality around the world in 2012.

![Figure (1.1): Distribution of global injury mortality by cause (WHO, 2012),](image)

Recently, traffic accidents and injuries increased dramatically in Palestine. In the period from 2007 to 2013, traffic accidents increased 2.5 times and the number of injuries increased twice. These escalations reflect the size of the problem and the need for new strategies to manage the traffic situation in Palestine (GPD, 2013).

1.2 Research Objectives

The main objectives of this research are:

- To collect and analyze traffic accidents database in the Gaza Strip between 2000 and 2013.
- To calibrate a model for predicting the existing of black spots using Artificial Neural Network (ANN).
1.3 Research Methodology

The objectives of this study will be achieved through performing the following tasks:

- Conduct a literature survey on previous studies of traffic accidents analysis, black spots analysis and prediction models. This review will cover factors that may affect accident repetition in Gaza Strip.
- Conduct a literature survey on the use of artificial neural networks (ANN) in Black Spot analysis.
- Collect data on traffic accidents in Gaza Strip, including accident location, time, damage, etc... These data will be use in traffic accident analysis and black spot identification and ranking process.
- Create a team that is consist of members from all governorates to locate traffic accidents data on a map of Gaza using ArcGIS program to study their spatial characteristics.
- Use data mining concepts to fill the missing data and to correct inconsistencies data on the databases.
- Determine black spot criteria in Gaza Strip and use it to identify black spot and to find the most dangerous locations in Gaza Strip.
- Rank traffic accident spots using adopted black spot criteria to create the spots database to use it in model creation and calibration.
- Determine the factors that are contribute on black spot occurrence in Gaza Strip.
- Use the data to create and calibrate a neural network model to predict and rank of black spots in Gaza Strip using Visual Gene Developer program (VGD).

1.4 Research Structure

The current study was divided into eight chapters as follows:

Chapter One is an introductory chapter that defines the problem statement, the objectives, the methodology and an overview of this study.

Chapter Two presents the definitions and history of traffic accidents, factors contributing to them, traffic accident classification and causalities and traffic safety measurements.

Chapter Three discusses data mining concepts, the functionalities of data mining spatial data mining and using data mining in black spot analysis.

Chapter Four explains the accident investigation and reduction definition, black spot identification and modeling and using GIS in black spot analysis.

Chapter Five deals with the fundamentals of Artificial Neural Network (ANN) showing their definition, the terminology used, as well as the advantages and disadvantages of them. The mechanism of ANN and the algorithms used for training them are also reviewed. Finally, it discusses some previous studies about using ANN in black spot modeling.
Chapter Six presents the data resources in Palestine and discusses it in terms of safety measurements. After that, the chapter discusses preprocessing the data step to prepare it to be used in the model.

Chapter Seven presents the black spot criteria in the research, spatial data preprocessing, and the ANN topology. After that, the chapter presents the model validation process and discusses the results.

Chapter Eight presents conclusions and recommendations for future work.
2 TRAFFIC ACCIDENTS ANALYSIS

This chapter contains definition of traffic accident and traffic accident history. After that, the chapter discusses the major exposure of traffic safety measurements.

2.1 Traffic accident definitions

There are many definitions of Traffic Accidents, In the United States it is defined as "An event that produces injury and/or damage, involves a motor vehicle in transport, and occurs on a traffic way or while the vehicle is still in motion after running off the traffic way" (IRTAD, 1997).

Another definition of traffic accident by Hanover Insurance Group is "unintended occurrence that caused or could have caused personal injury or property damage. Includes “near miss” accidents in which luck was the sole reason no one was hurt and/or nothing damaged" (HIG, 2006).

In Hungary, the definition of traffic accident is "any unexpected, unintentionally caused traffic event which results in fatality or personal injury or some material damage" (IRTAD, 1997).

The United Nations Economic Commission for Europe (UNECE) defines the accident as "Accidents which occurred or originated on a way or street open to public traffic; which resulted in one or more persons being killed or injured and in which at least one moving vehicle was involved" (IRTAD, 1997).

In Palestine; the Ministry of Transportation (MOT) defines road accident as an unintended occurrence that happened without any prepared plan by a vehicle, more than one vehicle, pedestrians, animals, or any other properties on the road, and usually causes damages which may be classified as a slight damage in properties or huge damage led to death or permanent disability (MOT, 2005).

Generally, Traffic Accidents Definition can be summarized in four major points as follow:

1. Unintended Occurrence.
2. Caused by a vehicle(s), vehicle loads, pedestrians, animals or any other properties (Some definitions require at least one moving motor vehicle).
3. In public areas, even that some definitions include private areas.
4. May cause damages ranging from slight damage in properties to huge damages which lead to death or permanent disability.

2.2 History of traffic accidents

Those who study the history of traffic accidents know that traffic accidents happened long time before automobile actually appeared on the roads. In the 16th century, Mother Shipton said "Carriages without horses shall go and accidents will fill the world with woe" (Shinar, 2007).

But the first fatal recorded automobile traffic accidents was in August 17, 1896, where a forty-four years old mother called Bridget Driscoll became the first road fatality in the world. She was hit by a car that was travelling at 4 MPH Speed (6.4 Km/hr) – Which was described as a tremendous speed according to witnesses-. At the
investigation, the officer said, "This must never happen again" (Shinar, 2007). In 1896, two fatal accidents were registered in the UK, and in 1889 one fatal accident was registered in USA, this was the beginning of the stream of death caused by traffic movement (Norman, 1962).

In 1957, the World Health Organization (WHO) recorded 102,552 persons where killed (out of 650 million inhabitants) by motor vehicle accidents in the 47 WHO member countries, and there is no records for the number of people who were injured or seriously disabled (Norman, 1962).

Nowadays; WHO mentioned that the number of fatal accidents has risen to 1.3 million accidents yearly, which cost more than 500 billion USA dollars. This cost represents more than 1% of the gross domestic product (GDP) (Michelin, 2013). The effect of the traffic accident exceeds that, Studies show that in 2004, road traffic accidents were considered as the ninth leading causes of death and it is forecasted to reach the fifth rank in 2030 (WHO, 2012).

These numbers reflect the needs of national plans to improve traffic safety around the world.

<table>
<thead>
<tr>
<th>2004 Rank</th>
<th>Disease or injury</th>
<th>2030 Rank</th>
<th>Disease or injury</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ischaemic heart diseases</td>
<td>1</td>
<td>Ischaemic heart diseases</td>
</tr>
<tr>
<td>2</td>
<td>Cerebrovascular diseases</td>
<td>2</td>
<td>Cerebrovascular diseases</td>
</tr>
<tr>
<td>3</td>
<td>Lower respiratory infection</td>
<td>3</td>
<td>Chronic obstructive pulmonary disease</td>
</tr>
<tr>
<td>4</td>
<td>Chronic obstructive pulmonary disease</td>
<td>4</td>
<td>Lower respiratory infection</td>
</tr>
<tr>
<td>5</td>
<td>Diarrhoeal diseases</td>
<td>5</td>
<td>Road Traffic injuries</td>
</tr>
<tr>
<td>6</td>
<td>HIV/AIDS</td>
<td>6</td>
<td>Trachea, bronchus, lung cancer</td>
</tr>
<tr>
<td>7</td>
<td>Tuberculosis</td>
<td>7</td>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td>8</td>
<td>Trachea, bronchus, lung cancer</td>
<td>8</td>
<td>Hypertensive heart disease</td>
</tr>
<tr>
<td>9</td>
<td>Road Traffic crushes</td>
<td>9</td>
<td>Stomach cancer</td>
</tr>
<tr>
<td>10</td>
<td>Prematurity and low birth weight</td>
<td>10</td>
<td>HIV/AIDS</td>
</tr>
</tbody>
</table>

It is important to mention that the first analysis of traffic accidents was published in the 30s, but they were limited in publishing and remained essentially prevented, just like Gilutz (1937) (Shinar, 2007).

2.3 Factors contributing to traffic accidents

The major factors contributing for traffic accidents could be summarized to: (PIARC, 2007), (O'Flaherty, 1997):

1. Human behavior factor (Road Users Mistakes): It is considered as the most important factor in traffic accidents because it contributes into more than 90% of traffic accidents. It includes drivers, passengers and pedestrians behaviors. It is influenced to many elements that relates to the individuals and their abilities, skills and experiences. These elements can be divided to four major elements; perceptual errors, lack of skill, manner of execution and impairments.

2. Vehicles factor: It is generally happened as a result of lack of maintenance by the vehicle user just like defective brakes and tires. Furthermore, it contributes to about 10% of traffic accidents.

3. Road factor: It includes the adverse road design, adverse environment, inadequate road furniture or marking and other unexpected obstructions. And it contributes to about 30% of traffic accidents.
Figure (2.1): Factors contributing to traffic accidents

(a) (PIARC, 2007), (b) (O'Flaherty, 1997).

Geurts et al. (2005) studied the reasons of concentration of traffic accidents in specific segments. They found that accidents in black spots depend on left-turns at signalized intersections, collisions with pedestrians, loss control of the vehicle and rainy weather conditions. On the other hand, accidents occurring outside black spots depend on left turns at intersections with traffic signs, head-on collisions and drunken road user(s).

2.4 Traffic accidents classifications

There are many types of Traffic Accidents classifications. This section is summarized the major types:

1. Classification due to type of damage:
   - Personal injury accident (PIA): It is an accident which involves an injury. It refers to the accident as the event (which may involves several vehicles and several casualties) (Slinn, 2005). PIA is categorized according to the most serious injury sustained by any person involved. For example, if any person is killed in the accident, the accident is classified as a fatal accident (TxDOT, 2008). The different definitions of the traffic accidents casualties will be discussed in the following section.
   - Property damage only (PDO); It is an accident which does not involve people who sustain personal injuries (O'Flaherty, 1997). In some countries, police authorities keep limited information about damage-only accidents (Slinn, 2005).

2. Classification due to the location of motor vehicle:

   There are many categories where the definition of traffic accident depend on its location just like:
   - Relative to junction (TxDOT, 2008):
     - At Intersection accident; It is an accident where the first harmful event occurs within intersection limits.
     - Intersection related accident; It is an accident where the first harmful event happened on an approach of intersection, however, it is a result
of an activity, behavior or control related to the traffic movements through the intersection.

- Driveway access accident; It is an accident where the first harmful event occurs on a driveway access.
- Non-intersection accident; It is any accident that is not included in the previous types.

- Land use character (TxDOT, 2008):
  - Urban accident; It is an accident happens at a location within city limits -more than 5,000 population-. 
  - Rural accident is an accident which cannot be classified as an urban accident.

3. Motor vehicle accident classification

- Non-collision accident; it is any accident involving a motor vehicle in moving, that may occur without collision and it can be divided into (TxDOT, 2008):
  - Overturning accident; where a vehicle in moving overturns for any reason without any collusion.
  - Other non-collision accident; any accident other than overturning and collision; just like explosion of any part of the vehicle.

- Collision accident; where accident between a motor vehicle, its loads, its parts or object set in motion by the motor vehicle colloid with other things just as (TxDOT, 2008):
  - Pedestrian.  
  - Motor vehicle in transport.  
  - Parked motor vehicle.  
  - Railway train.  
  - Pedal cyclist.  
  - Animal.  
  - Fixed object.  
  - Other object just like carts and fallen trees.

2.5 Traffic accidents casualties

For each country, there is a different definition of the traffic accident casualty. In this section, some of these definitions will be discussed;

UNECE (United Nations Economic Commission for Europe) definition may considered as the most common definition of traffic accident casualties, which is divided into three types:

1. Killed casualty (fatal casualty): it is defined as any person dies during an accident or within 30 days as a result of an accident (UNECE, 2009).

The following points illustrate some other definitions of the fatal accidents in some different countries (IRTAD, 1997), (UN, 2007), (Issa, 2008);

- In Portugal, it is defined as the casualty who dies at the scene of the accident, during transport from the scene of the accident or immediately after it.
• In Turkey and Spain, it is defined as any person dies during an accident or within 30 days as a result of an accident.
• In Korea, it is defined as any person dies during an accident or within 3 days as a result of an accident.
• In France, it is defined as any person dies during an accident or within 6 days as a result of an accident.
• In Italy and Latvia, it is defined as any person dies during an accident or within 7 days as a result of an accident.

2. Serious casualty: it is defined as any injured person who was hospitalized for more than 24 hours (UNECE, 2009). This definition depends on hospitalization process for a period of time to define the serious casualty.

Some countries have the same definition concept but with different periods, while other countries have other characteristics that control the definition. Some of these definitions are illustrated in the following points (IRTAD, 1997), (UN, 2007), (DfT, 2008):
• Australia, Austria, UK, Sweden and other countries required the hospitalization process without time limit to define serious casualty.
• France required the hospitalization process for more than six days in serious casualty definition, while Poland required seven days.
• Czech, Denmark, Japan, Italy and others do not require hospitalization in serious casualty definition.
• Sweden, Czech, Denmark, the UK, Poland and others use the type of injury in serious casualty definition.
• Norway and Switzerland use the disability to work to define serious casualty.
• New Zealand, the UK, Ireland and Japan require medical treatment in serious casualty definition.

3. Slight casualty: it is defined as any injured person excluding seriously injured persons (UNECE, 2009).

As shown in serious casualty, some countries have the same definition concept, while other countries have other characteristic that control slight casualty definition. Some of these definitions is illustrated in the following points (IRTAD, 1997), (CSO, 2010), (DfT, 2008):
• Germany, Mauritius, Denmark, Czech, Sweden and other countries adopt UNECE definition of slight casualty.
• Netherlands, Portugal, Switzerland, define it as non-hospitalized casualty, while France define it as a casualty who hospitalized for less than 6 days and Canada define it as a casualty who hospitalized for less than 24.
• UK, Poland, Switzerland, Greece and others use the type of injury in slight casualty definition.
• Portugal, Greece, France and Japan require medical treatment in the
definition of slight casualty definition.

In the USA, the severity of traffic accident casualties is classified depending on the
disability to work. It is divided into 5 severities as follows (IRTAD, 1997), (TxDOT, 2008):

1. Fatal injury: it is defined as any injury that result death within 30 days as a result of an accident.
2. Incapacitating Injury: it is defined as any injury, other than a fatal injury, which prevents the casualty from living his/her normal life. Just like walking, driving or doing any activity he was capable of performing before the accident. This type of causalities is defined as serious casualty in national comparison studies.
3. Non-incapacitating evident Injury: Any injury, other than a fatal casualty or an incapacitating casualty, which is evident to observers at the scene of the accident in which the injury occurred. This type of casualities is defined as slight casualty in national comparison studies.
4. Possible Injury: Any Injury reported or claimed which is not a fatal injury, incapacitating injury or non-incapacitating evident injury. This type of causalities is defined as slight casualty in national comparison studies.
5. No injury: in this situation, there is no reason to believe that the person received any bodily in the traffic accident.

In Palestine, the accidents severity depends on the type of injury; and it is divided into four major types (MOT, 2005):

1. Killed casualty (fatal casualty): Any person killed during an accident or within 30 days as a result of an accident.
2. Serious casualty: Any casualty that has major incapacitating injury, just like skull, backbone or chest injuries, abdominal injuries or internal bleeding.
3. Moderate casualty: Any casualty that has minor incapacitating injury, just like severe lacerations and broken or distorted limbs.
4. Slight casualty: Any casualty that has no incapacitating injury, just like contusions and bruises and superficial wounds.

### 2.6 Traffic safety measurements

One of the major goals of traffic management is to decrease the number of accidents on the road. However, in order to evaluate the traffic safety trends, it is not realistic to use the absolute number of crashes; because it is expected to increase over the time as the number of cars and drivers increase. So trends in road safety are usually measured in terms of the rate of crashes or injuries; by dividing the number of crashes or injuries by a measure of exposure. There are several different rates which used to track the trends of safety over time. Each one uses a different exposure and gives a different measure of risk (Shinar 2007). Table (2.2) gives some different exposures used in traffic safety measurements.
Table (2.2): Different exposures used in the safety measurements (Shinar 2007).

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Exposure Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk per population</td>
<td>• It is calculated by dividing the number of risk representative by the size of the population.</td>
</tr>
<tr>
<td></td>
<td>• It is a simple measure.</td>
</tr>
<tr>
<td></td>
<td>• It is a wide use measure &quot;available in most countries&quot;.</td>
</tr>
<tr>
<td></td>
<td>• It gives the average risk per person.</td>
</tr>
<tr>
<td>Risk per drivers</td>
<td>• It is calculated by dividing the number of risk representative by the number of licensed drivers in the population.</td>
</tr>
<tr>
<td></td>
<td>• It is considered as a limited measure of travel exposure because not all drivers have cars.</td>
</tr>
<tr>
<td>Risk per motor vehicles</td>
<td>• It is calculated by dividing the number of risk representative by the number of motor vehicles.</td>
</tr>
<tr>
<td></td>
<td>• It is considered as a limited measure of travel exposure because it omits non-motorized transport. But in most countries, the definition of the traffic accident must involve a motor vehicle.</td>
</tr>
<tr>
<td>Risk per the total number of miles or kilometers driven</td>
<td>• It is calculated by dividing the number of risk representative by the total number of miles or kilometers driven.</td>
</tr>
<tr>
<td></td>
<td>• It is considered as a realistic measure because only vehicles that are actually moving on the road are involved.</td>
</tr>
<tr>
<td></td>
<td>• It is greatly affected by the level of urbanization and demographic characteristics of the road users.</td>
</tr>
</tbody>
</table>

With these potential denominators and at least three qualitatively different numerators - number of accidents, number of injured, and number of fatalities - twelve different indices can describe the state of traffic safety in any country. The following measures are very common measures and they were adopted by WHO and with their permission (Shinar 2007):

1. Number of injuries and number of fatalities: the first one is an absolute figure indicating the number of people injured in road traffic crashes. The other one is an absolute figure indicating the number of people who dies as a result of road traffic crashes. These measures are useful for planning at the local level for emergency medical services and for calculating the cost of medical care. On the other hand, it is not very useful to make comparisons and it may have some error due to the large proportion of not reported slight injuries.

2. Fatalities per 100,000 populations: it is a relative figure represents the ratio of fatalities to the population. This measure shows the impact of road traffic crashes on human population and it is useful for estimating severity of crashes.

3. Fatalities per 10,000 vehicles: it is a relative measurement represents the ratio of fatalities to motor vehicles. This measure is limited because it omits non-motorized transport –note that in most countries the definition of the traffic accident must involve a motor vehicle-.

4. Fatalities per vehicle-km traveled: it is a relative figure that represents the ratio of fatalities to kilometer traveled. This measure is useful for international comparisons. However, it omits non-motorized transport.
3 DATA MINING

This chapter discusses the data mining concept to convert data to knowledge. At the beginning it discusses the knowledge hierarchy (DIKW hierarchy). After that it discusses the definition of data mining and its functions. Finally, it talks about the data mining applications in traffic accidents and black spots analysis.

3.1 DIKW hierarchy

The data–information–knowledge–wisdom (DIKW) hierarchy, referred to variously as the ‘Knowledge Hierarchy’, the ‘Information Hierarchy’ and the ‘Knowledge Pyramid’ is one of the fundamental (Rowley, 2007).

There are many definitions of these terms; Ackoff (2009) defines as follow:

- The data is the most basic level of the hierarchy which consists of factual information. In other words it is raw and unprocessed set of observations and measurements.

- Information is the level of structured data which can answer some of the basics questions (e.g. what, when, where and who). The structural system enables the relationships and connections analysis across the data.

- Knowledge is the level of understanding the process and makes it possible to transform information into instructions. Knowledge can be obtained either by transmission from another who has it, by instruction, or by extracting it from experience.

- Wisdom is the level that answers the question “why”. This level incurs a sort of personal interpretation of the knowledge which requires a mental function that we call judgment. It is affected by the ethical and aesthetic values that are inherent to the actor and are unique and personal.

Figure (4.1) shows the relations in DIKW hierarchy.
3.2 Definition of data mining

Data mining is an interdisciplinary exercise, where statistics, database technology, machine learning, pattern recognition, artificial intelligence, and visualization all play a role. As it is difficult to define sharp boundaries between these disciplines, it is difficult to define sharp boundaries between each of them and data mining. At the boundaries, one person's data mining is another's statistics, database, or machine learning problem (Hand et al., 2001).

Generally, many researchers define data mining as a synonym for the term “knowledge discovery from data (KDD)”, while others define it as a step in knowledge discovery process (Han et.al, 2012).

A definition of data mining is the process of extracting hidden predictive information from large databases. This process helps decision makers to invest the databases to choose the optimum decision. This process is supported by three major technology 1) massive data collection, 2) powerful multiprocessor computers and 3) data mining algorithms (Krishnaveni, 2011).

Another definition is the process of selection, exploration, and modeling of large data to discover new regularities or new relations to get clear and useful results (Giudici, 2003).

3.3 Knowledge Discovery from Data steps (KDD steps)

The KDD process is a repeated sequence of the following steps (Han et.al, 2012):

- Data preprocessing: the importance of this step is to improve the data quality to make it satisfy the requirements of the intended use. The data quality comprises of different factors including accuracy (no errors and no values that deviate from the expected), completeness (no missed attribute), consistency (do not containing discrepancies in the department codes used to categorize items), timeliness (use the data before the end of the time period), believability (how much the data are trusted by users?), and interpretability (how it is easy to understand the data?). The data preprocessing can be divided to the following steps;
  - Data cleaning: it is routines attempt to fill in missing values (by ignoring the tuple, filling in the missing value manually, using the most probable value to fill in the missing value, etc…), smooth out noise while identifying outliers (smooth a sorted data value by using the values around it –binning-, using regression methods, excluding the outlier values, etc…), and correct inconsistencies in the data (using external references, use integrating techniques to make standard terminology, etc…).
  - Data integration: it is the process of merging data from multiple data sources. Data integration techniques used to reduce and avoid redundancies and inconsistencies in the new data set, and it is also used to detect the tuple duplications – that define as where there are two or more identical tuples for a given unique data entry case-. Statistical correlation tests are used to detect some kind of redundancies. Data cleaning process is needed after data integration to resolve data value conflict.
Data selection (reduction): a set of techniques that can be applied to reach a reduced set of data which is smaller than the original data, but very close and can be used to represent it. There are many strategies that are used to reduce data (e.g. dimensionality reduction - reducing the number of random variables or attributes under consideration-, numerosity reduction – techniques to replace the original data volume by alternative set of data just like regression, log-linear models and histograms- and data compression – reconstructing the data to create new lossless or lossy set of data-).

Data transformation: routines convert the data into appropriate forms for mining, just like removing the noise data (smoothing), constructing and adding new attribute from the given set of attributes to help the mining process (attribute construction), applying summary operations to the data (aggregation), scaling the data to fall in small ranges (normalization) and replacing the values of numeric attribute by interval labels or conceptual labels (discretization).

- Data mining: it is essential process where intelligent methods are applied to extract data patterns. It will be discussed in the following section.
- Pattern evaluation: it is important to identify if the pattern is interesting pattern or not. The properties of interesting pattern is (1) easily understood by humans, (2) valid with new set of data with degree of certainty, (3) potentially useful, and (4) novel. It is unrealistic and inefficient to generate all possible patterns. So it is more efficient to put experts’ constraints to focus in the search. In other hand, generating only the interesting patterns is considered as an optimization problem.
- Knowledge presentation: it is the techniques of visualize and present the mined knowledge to users (e.g. decision makers).

3.4 Data mining functionalities
Data mining functionalities are used to specify the kinds of patterns to be found in data mining tasks. Generally, data mining tasks divided into two major categories (Han et.al, 2012);

1. Descriptive tasks: those describe the target data properties. It includes;
   - Characterization: is a summarization of the general characteristics or features of a target class of data. The output of data characterization can be presented in various forms just like pie charts, bar charts, curves, multidimensional data cubes and multidimensional tables or presented as general relations or rules form which called (characteristic rules).
   - Data discrimination: is a comparison of the general features of the target class data against the general features of one or multiple other classes which are defied by the user. The output presentations are similar to those for characteristic descriptions. However, discrimination description should include comparative measures that help to distinguish between the target and other classes which called (discriminant rules).
   - Frequent patterns: are patterns that occur frequently in data and include the frequent items, frequent subsequences and frequent substructures.
Mining frequent patterns leads to the discovery of interesting associations and correlations within data. It is important to identify interesting rules that have minimum support and minimum confidence.

- **Cluster Analysis**: is a process used to generate class labels for a group of data based on the principle of maximizing the intraclass similarity and minimizing the interclass similarity (objects within a cluster have high similarity in comparison to one another, but are rather dissimilar to objects in other clusters). It facilitates taxonomy formation that classified observations in hierarchy classes depending on their similarity.

- **Outlier Analysis**: outliers are the objects that do not comply with the general behavior or model of the data. They may be detected using statistical tests (assuming a distribution or probability model for the data), using distance measures (objects that are remote from any other cluster are considered outliers) or using density-based methods (outliers in a local region although they look normal from a statistical point of view).

2. **Predictive tasks**: those use the properties of the current data to make predictions. It includes:

- **Classification**: is the process of creating a model or a function that describes data classes or concepts based on training data. The model is used to predict the class label of objects for which the class label is unknown. The model may be represented as classification rules, mathematical formulae, decision tree (flowchart like tree structure, where each node denotes a test on an attribute value, each branch represents an outcome of the test, and tree leaves represent classes or class distributions), artificial neural network (will discussed in chapter five) etc...

- **Regression**: is a statistical methodology that is often used for numeric prediction. It also encompasses the identification of distribution trends based on the available data.

### 3.5 Mining spatial and spatiotemporal data

In the real live, it is important to discover knowledge from spatial and spatiotemporal data. Spatial data includes geospace-related data stored in geospatial data warehouse and it can be in vector format, raster format, or in the form of imagery and geo-referenced multimedia. Spatiotemporal data are data that relate to both space and time (e.g. weather, earthquakes, hurricanes and moving-object) (Han et.al, 2012). Geographic information system (GIS) enable capturing, storing, analyzing, and managing data and associated attributes which are spatially referenced to spatial location (Lavrač, et. al. 2008).

Some major topics of spatial data mining are mining spatial associations and co-location patterns, spatial clustering, spatial classification, spatial modeling, and spatial trend and outlier analysis (Han et.al, 2012).

While spatiotemporal data mining include moving-object data, mining periodic patterns for one or a set of moving objects, and mining trajectory patterns, clusters, models, and outliers (Han et.al, 2012).
3.6 Data mining in traffic accidents analysis

Various studies have addressed the different aspects in traffic accidents analysis, with most focusing on predicting or establishing the critical factors influencing injury severity (Ro, et al. 2005).

Chong et al. (2005) made a data mining research focusing on building tree-based models to analyze freeway accident frequency. Using two years (2000-2001) accident data of National Freeway 1 in Taiwan, they developed classification and regression tree (CART) and negative binomial regression models to find the relations between traffic accidents and highway geometric variables, traffic characteristics, and environmental factors. They found that the ADT and precipitation variables were the main factors in freeway accident frequency.

Sohn et al. (2002) used various algorithms to improve the accuracy of individual classifiers for two road accident severity categories. Using ANN and decision tree individual classifiers, three different approaches were applied: classifier fusion based on the Dempster–Shafer algorithm, the Bayesian procedure, and logistic model. They found that a clustering-based classification algorithm is the most useful procedure to classify road accident in Korea.

Ng et al. (2002) used a combination of cluster analysis, regression analysis, and number of traffic accidents to assess road accidents risk in Hong Kong. They developed an algorithm which improved accident risk estimation compared to estimates based on historical accident records alone especially for fatality and pedestrian related accident analysis. The researchers claimed that the algorithm could improve road safety in planning process and could be used to identify areas with high accident risk.

Sohn et al. (2001) tried to recognize patterns of road traffic accidents severity in Korea. They observed that an accurately estimated classification model for several road traffic accidents severity give practical information for accident prevention. In their research, they used three data mining techniques (neural network, logistic regression, and decision tree) to select a set of influential factors and to construct classification models for accident severity. The three approaches were then compared in terms of classification accuracy. They found that accuracy did not differ significantly for each model, and that the protective device was the most important factor in the accident severity variation.

Researchers discussed black spot models, GIS models and ANN models will be discussed in the following chapters.
4 BLACK SPOT ANALYSIS

This chapter discusses the concept of accident investigation and reduction and the definition of black spots. After that the chapter discusses black spot modeling using different techniques.

4.1 Accident investigation and reduction

Accident investigation is considered as the first step in traffic accident reduction process, it provides the researchers and decision makers with the necessary information to improve the current situation and to find the optimum solutions with low-cost engineering programs. It includes data collection, storage and retrieval. The second step is identifying the hazardous locations, then diagnosis of the accident problem and finally the selection of plans to be implemented in government plans (O'Flaherty, 1997).

To achieve acceptable reduction in traffic accidents, it is important to ensure the type and quality of the traffic accident data collected in accident investigation process. The minimum data that can provide road engineers with the necessary information are accident identification, time, location, accident type, vehicles involved and accident consequences (PIARC, 2007).

In traffic accident reductions process, there are four major strategies which can be implemented to achieve reasonable reduction in traffic accidents. The data availability and the fund availability are considered as major factors in strategy selection. These strategies are concluded as follow (O'Flaherty, 1997):

- **Single site approach**: this strategy is probably the most common approach in accident reduction strategies. It involves addressing a specific site with 1) a much higher than average concentration of particular type accidents, 2) in a specific period generally 3 years. Specific site includes individual intersection, short length of road and small area. In single site approach, a large accident reduction is predicted (typically 33%).

- **Mass action approach**: this strategy is considered as the second common approach in accident reduction strategies. It involves addressing all locations having a similar accident problem over the under consideration area with a satisfy reduction measurements. An example of the target location is all T-intersections in the understudying area. In mass action approach, 15% average accident reduction is predicted.

- **Route action approach**: this strategy is used when there is one or more roads, with the same characteristics, have a significant increment of accidents than others. In this strategy the road is generally divided into sections from 0.5 to 1 km. In route action approach, 15% accident reduction is predicted in the first year.

- **Area action programs**: this strategy is used when there is a significant proportion of accidents that separately scattered in under consideration urban area. It involves addressing major problems over the traffic network in one part of the area. In area action approach, 10% accident reduction is predicted in the first year.
4.2 Black spot identification, ranking and selection

In traffic safety, there is no specific definition of black spot in road networks. Generally, the identification of black spot is divided into two major approaches depending on the type of accident data used in identification process, as follows:

- Identification of black spot depending on historical accident data

It is the traditional way to identify black spot and it is considered as the most common identification method of black spots. It is used in too many countries in the world such as the USA (O’Flaherty, 1997), the UK (O’Flaherty, 1997), Australia (Geurts et.al, 2003), Jordan (Alhrab, 2001), Poland (Szczuraszek, 2000) and others. Generally; black spot is defined as the locations which have a higher than average crash rate (accident per vehicle) or crush frequency (accident per year or accident per kilometer) (Geurts et.al, 2003). But each country has its own criteria in black spot identification as shown below;

In the UK, there is no specific procedure to identify black spots. A simple identification of black spot is any section, spot or area where the number of injuries is more than a specific number of injuries. After that, the selected sections may be sorted by its numeric identification in descending order from the highest to the lowest. Some authorities in UK identify the black spot by the accidents exposures just like crash rate or crash frequency (O’Flaherty, 1997).

In the USA, there is a common simple method to identify black spots. It is depending on weighting of accidents in a specific location depending on its severity. The weighting factor of each severity is depending on the relative cost of the accidents type (O’Flaherty, 1997).

In Australia, the definition of black spot depending on the number of the injury accidents in three years and the average annual daily traffic in the under consideration area. The black spot is any location that have three or more injury accidents and give a value of risk coefficient \( R_k \) greater than or equal to 0.8, where;

\[
R_k = \frac{\text{Number of Injury accidents in 3 years}}{0.5 + 7 \times 10^5 \times \text{AADT}}
\]

A 250 m slide window is used in the identification process. And it is important to illustrate that the identification process does not depend on accident severity (Elvik, 2008), (Geurts et.al. 2003).

In Flanders, black spot is any sliding window with 100 m maximum length, where three accidents recorded in police reports in three sequential years (Elvik, 2008). The accident ranking depends on accident severity where;

\[ S = \text{No. slight accidents} + 3 \times \text{No. serious accidents} + 5 \times \text{No. fatal accidents}. \]

In Germany, there are two procedures to define black spots depending on the length of the identification period. If the length of the identification period is 1 year, the black spot is defined as any location where five accidents of a similar types happened (irrespective of its severity). However, if three years identification period is used, the black spot is defined as any location where five or more injury accidents have been recorded or any location where three or more serious accidents have been recorded. The identified location is selected by maps showing plots of accidents and does not exceed 100 m length (Elvik, 2008).
In Hungary, the black spot is defined as any location where four accidents recorded in police reports in three sequential years. A 1000 m sliding window is used in the identification process outside built-up areas while a 100 m sliding window is used in the identification process inside built-up areas (Elvik, 2008).

In Portugal, there are two procedures to identify black spot; the first one is defined black spot as any location, determine by a 200 meter sliding window or less, where 5 accidents happened in one year with severity index more than 20. The severity index is calculated by the following equation;

\[ SI = \text{No. slight accidents} + 10 \times \text{No. serious accidents} + 100 \times \text{No. fatal accidents}. \]

The second procedure depends on the expected number of accidents, as will be shown in the next section (Elvik, 2008).

Some researchers used this approach -historical accident data- to identify and rank black spots. Birjs et.al. (2006) made a multivariate Poisson distribution model to identify black spots, they used the number of accidents in each site and the type of causality to develop a cost function that ranks sites and identifies black spots.

- **Identification of black spot depending on the expected number of accidents**

It is considered as the state of art approach to identify black spots. It is depends on analyzing each site location to identify risk factors for accidents after that multivariate accident prediction model should be developed (Elvik, 2007). In addition, it is considered as the best approach from theoretical point of view (Sørensen, 2007).

This approach is used by national and regional public roads administrations using comprehensive and interoperable traffic and accidents data (Sørensen, 2007).

As mentioned before, Portugal is one of the countries that have a definition of black spots depending on the expected number of accident. This procedure is tested in 1998 and applied in Portugal since 2004. A unique accident prediction model is fitted for each class of roads depending on number of carriageway, carriageway width and number of lanes in each direction. All models were developed for non-intersection accidents using the following general equation (Elvik, 2007):

\[ E(\lambda) = \beta_1 \times ADT^{\beta_2} \times CW^{\beta_3} \times e^{\left(\frac{\beta_4 \times ADT}{1000}\right)} \]

Where \( E(\lambda) \) is the expected number of accident in 5 years period, ADT is the average daily traffic, CW is carriageway width and \( \beta_i \) are parameters estimated by model fitting. The worst 1/1000 or 2/1000 cases are selected for diagnoses and intervention.

In Denmark, this procedure is used to identify black spots. The road network is divided into three partitions a) Road Section b) Roundabout and c) Intersections. Each partition is divided into sub-group which called ap-group. For each ap-group, a specific model is used to identify and rank black spots. The road section model is generally defined by using the formula;

\[ \mu = a \times AADT^b \]

where \( \mu \) is the expected number of accidents, AADT is the average annual daily traffic and other symbols are parameter estimated for each ap-group.

While intersections formula is;
\[ \mu = a \cdot AADT_{mA}^{b1} \cdot AADT_{mI}^{b2}. \]

where \( AADT_{mA}^{b1} \) and \( AADT_{mI}^{b2} \) are AADT for major and minor roads respectively (Vistisen, 2002).

Recent studies adopted this approach in identifying black spots; just like Mustakim et al. (2008), Chong et al. (2004), Akiyama et al. (2000) and others.

4.3 Black spot modeling

Black spot modeling is used to estimate network safety in existing national and regional public roads and ensure new roads safety before implementation.

Recent studies work to identify main factors that effect on traffic safety and create a formula that is fit for each case. Bhat et. al. (2013) tested 19 factors to find those can represent the actual situation. They found the relationship of each individual factor with the dependent variable and filtered them using statistical parameters that check the goodness of fit just like coefficient of determination (\( R^2 \)) and Se/Sy factor (residual standard deviation/ sample standard deviation). After that, they omit weak factors and create new model strong factors. The final formula is:

\[ y = 6.037 + 0.041CW^2 + 0.471e^{ST} + 4.346T + 2.17LU - 0.338SC \]

where CWis carriageway width, ST is shoulder type, T percent of trucks in the traffic stream, LU represents land use and SC represents speed component which can be calculated as a function of road condition.

Rokade et. al. (2010) developed a prediction model in India using the following factors 1) Road cross-section dimensions, 2) Traffic volume, 3) Speed, 4) Road shoulder width, 5) Lighting conditions, 6) Traffic Signs, 7) Traffic Signals. The coefficient of determination (\( R^2 \)) for model that created is 0.9999 (and that mean the ratio of explained the variance to the total variance is 0.9999). The final model formula is as follows;

\[ Y = a_0 + \sum_{i=1}^{n} a_iX_i \]

where \( Y \) is weight of the spot from 0% (high risk) to 100% (no risk), and it is calculated by the following formula Total weight = \( \sum_{\text{Individual Weights}} \frac{\text{Individual Weights}}{\text{maximum weight}} \times 100\% \), a\( _0 \) and a\( _i \) are factors and X\( _i \) are the factors affecting occurrence of accidents.

Oyedepo et. al. (2010) used multiple linear regressions to predict the natural of accident depending on the number and the type of causality. The accident is considered minor accident if the factor (F) equals 1. And it is considered as a serious accident if it equals 2 and it is considered as a fatal accident if it equals 2. F is calculated as follow;

\[ F = 0.119X_1 + 0.075199X_2 - 0.007916X_3 + 1.562 \]

where X\( _1 \) is the number of people killed in the accident, X\( _2 \) is the number of people injured and X\( _3 \) is the number of people involved in the accident. It is noted that if the number of people involved in the accident increases the nature of the accident is improved.

Rakha et.al. (2009) also used a linear regression model to predict traffic accidents with two major factors 1) AADT 2) distance to the first access road. They found that as
AADT increase as traffic accident increase. In addition, they recommended increasing the minimum spacing between access roads in freeways from 90 m to 180 m.

Wier et al. (2008) created a vehicle-pedestrian injury collision regression model in area-level based on environmental and population data in 176 San Francisco. The model includes many factors such as 1) traffic volume, 2) arterial streets without transit, 3) land area, 4) proportion of land area, 5) zoned for neighborhood commercial and residential-neighborhood commercial uses, 6) employee and resident population, 7) proportion of people living in poverty and 8) proportion aged 65 and older. The model explained approximately 72% of the systematic variation of vehicle-pedestrian injury collisions in census-tract. The general model form used for analyses is:

$$\ln(PIC) = \beta_0 + \sum \beta_i X_i$$

where PIC is the predicted vehicle-pedestrian injury collisions per census tract, $\beta_0$ and $\beta_i$ are factors and $X_i$ is the census-tract level data for predictor variable $i$.

Mustakim et al. (2008) created different models using multiple liner regression for rural roadway; they found that the main factors that effect on the model were 1) number of access points per kilometer 2) hourly traffic volume 3) time headway and 4) 85th percentile speed.

Mustakim et al. (2006) created a prediction model to find the black spots at the Federal route FT50 Batu Pahat-Air Hitam in Malaysia. They defined the black spot using weight functions, which are related to the average accident cost of each severity level. They used regression model with three factors to identify black spot, using the following formula;

$$\ln(APW) = (1.1640952 \log(AP) + 0.001727(HTV^{0.75} + V_{p85}^{1.25}))^2$$

where $APW$ is the accident point weightage, $AP$ is the number of access point per kilometer, $HTV$ is the hourly traffic volume and $V_{p85}$ is the 85th percentile speed.

Greibe (2003) also put accident prediction model for roads link and for intersections in urban area. He used many variables just as traffic flow, speed limits, number of lanes, lanes width, road facilities, central islands, land use and number of arms. For road links he used the following formula

$$E(\mu) = a. N^p \exp \sum B_j x_{ij}$$

where $E(\mu)$ is the expected number of accidents, $N$ is the AADT, $x$ is a factor depending on road geometry and environment of the road. While the others are estimated parameters.

And for junction he use the following formula

$$E(\mu) = a. N_{pri}^{p1} N_{sec}^{p2} \exp \sum B_j x_{ij}$$

where $N_{pri}$ and $N_{sec}$ are the AADT for primary and secondary direction respectively.
Turner et al. (1998) compared between generalized linear models in intersections. They found that the models that used conflict flow to identify black spot is better than other models that used approach flow.

4.4 Using Geographic Information System (GIS) in black spot analysis

A Geographic Information System (GIS) is a computerized system that is used for capturing, storing, querying, analyzing and displaying geographic data. GIS can be divided into two major software technologies i.e., Data Base Management System (DBMS) and Computer Aided Design (CAD). In addition, it is specialized to manage and analyze spatial data that referenced to a geographical location (Mandloi et al., 2003).

At recent time, it is very common to use GIS in black spot analysis and modeling. Isen et al. (2013) evaluated the black spots using GIS. They used three years data to identify the black spots (2009-2011) in Alappuzha, India. They started by collecting data from the field and State Crime Records Bureau. Then they analyzed the secondary data (accident data) using Weighted Severity Index (WSI) where;

\[ WSI = 41 \times \text{killed persons} + 4 \times \text{grievous injuries} + 1 \times \text{minor injuries} \]

Then they made a road inventory survey to measure the roadway geometric parameters like the roadway width, footpath width, median, shoulders, surface type, surface condition, edge obstruction, road markings, road signs, drainage facilities and adjoining land use. After that, they adopted a traffic study that is concerned with finding the traffic volume, speed and delay in the understudy area.

The methods that they suggested to improve safety is included improving the infrastructure facilities of the region just like increasing the number of the lanes, providing a median, providing a footpath, providing adequate drainage system and others.

Mandloi et al. (2003) made a GIS model to analyze the black spot. The model considered factors that affect the black spot just like road width, number of lanes, approximate number of vehicles per day, the radius of horizontal curve, type of road, drainage facilities, surface condition of the pavement, frequent vehicle type and presence of shoulders, edge obstructions, median barriers and ribbon development. They found that the model need very less additional data to use than road network map, so it is very easy to implement it in planning road safety measures.

Liang et al. (2005) made a traffic accident application to manage accident database entries using GIS. They collected and digitized accident data, and then they identified and ranked the black spots using several types such as ranking by accident cost, nodal analysis and analysis on link accidents. After that in-depth diagnosis of selected black spots was carried out to capture the near misses, approach speeds, vehicles and pedestrian flows and their maneuvers. The final step was to create a traffic accident application which called GIS-RAVS (Geographic Information System and Road Accident View System). The core function of the new application is to entering, searching and retrieving information from database while the main supportive function is to do node analysis and accident distribution plot and the minor function is to make a searching engine and accident ranking system.

Saccomanno et al. (2001) created a model depending on state of art approach of identifying black spots using accident data collected by Ontario ministry of transportation in Canada. The main features of the GIS model are 1) creating an
integrated Relational Data Base Management System (RDBMS) with the ability to query and create relevant statistics, 2) creating an accident prediction and analysis module applied to different route locations, and 3) creating a GIS platform for visual display of spatial analysis.

They used a prediction model developed by Nassar (1996) and Nassar et. al. (1994) using the data from the same area. The model was as follows:

\[
E(m)_i = ADTL^{2.242} \cdot LEN^{-0.696} \cdot e^{(0.1955 \cdot LN - 0.1775 \cdot SHW + 0.2716 \cdot MT2 + 0.5669 \cdot TS - 0.1208 \cdot PTC - 0.0918 \cdot Y91)}
\]

where \(E(m)_i\) = expected accident frequencies on road section, ADTL is the annual average daily traffic (AADT) per lane, LEN is the length of road section, LN is the number of lanes on road section, SHW represents the shoulder width of road section, MT2 represents the median type, TS represents traffic signal on road section, PTC represents pattern type commuter on road section and Y91 represents the data of year 1991.

The used data are: 1) Digital Cartographic Reference Base (DCRB), 2) Accident Data System (ADS), 3) Highway Inventory Management System (HIMS), and 4) Traffic Volume Inventory System (TVIS).

As mentioned before, the model predicts accidents by using the state of art approach, so it can be used to evaluate the effectiveness of alternative safety countermeasures designed to reduce accidents at unsafe locations or routes.

Eljamassi (2007) used GIS to locate the sites where the accidents happened frequently in Gaza (Palestine) and Grenoble (France). He used Geocoding to locate accidents into ArcGIS. Geocoding depends on the accidents postal address which is not available in Gaza, so he used the nearest milestone that is recorded in police report (like a restaurant or intersection). The accident data was collected during three years in France (1999-2003) and five year in Gaza (1998–2002). After that, he used spatial analysis tool to analyze 1) cartographic of total accidents and pedestrian accidents, and 2) cartographic of accident density and identifying the location that recorded high number of accidents.

In this research, the GIS will be used to prepare the data in preprocessing step and join it with spatial references to use the data to create a prediction model using artificial neural networks.

4.5 Effectiveness of black spot models

In black spot modeling, it is important to evaluate the effectiveness and the performance of the models, so it is important to create measurable indicators by sorting the black spots that is used in identification process to four major categories (Elvik, 2007):

1. Correct Positive: where both the expected number and the recorded number of accidents exceed the critical value of black spot definition.
2. False Positive: where the expected number of accidents does not exceed the critical value of black spot definition, while the recorded number of accidents exceeds.
3. Correct Negative: where both expected number and recorded number of accidents do not exceed the critical value of black spot definition.
4. False Negative: where the expected number of accidents exceeds the critical value of black spot definition, while the recorded number of accidents does not.

These categories are used to assess the model performance quantitatively; two of the most common diagnostic tests are sensitivity (the ratio between the number of correct positives to the number of total positives) and specificity (the ratio between the number of correct negatives to the number of total negatives) (Elvik, 2007).
5 ARTIFICIAL NEURAL NETWORK (ANN)

This chapter discussed the ANN definition and also discussed the major terminologies used in it. After that it talked about ANN mechanism and types. Finally, it is summarized some researched used ANN in traffic safety issues.

5.1 Definition of ANNs

An artificial neural network is a network of a large number of connected processing units, the so-called nodes or neurons. The neurons are connected by unidirectional communication connections. The strength of the connections between the neurons is represented by numerical weights. Knowledge is stored in the form of a collection of weights. Each neuron has an activation value that is a function of the sum of inputs received from other nodes through the weighted connections (An-Najjar, 2005).

5.2 Terminology used in ANNs

The following terminologies are the major terms in ANN;

Neuron (artificial): It has inputs from other neurons, with each of which is associated a weight - that is, a number which indicates the degree of importance which this neuron attaches to that input. It also has an activation function, and a bias. It is the processing element in ANN and they are called nodes also (An-Najjar, 2005).

Figure (5.1) shows a typical structure for ANN.

![Typical Structure for ANN](image)

\[ \sigma = f(Wp + b) \]

Where...

\[ \sigma \] = number of elements in input vector

Figure (5.1): Typical Structure for ANN (An-Najjar, 2005)
Weight of a neuron: it is determines how much notice the neuron pays to the activation it receives from previous neuron (Wilson, 2012).

Input units are a neuron with no input connections of its own.

Output unit is a neuron with no output connections of its own. Its activation thus serves as one of the output values of the neural net (Wilson, 2012).

Bias: the trainable weight for hidden neuron or output neuron which always has an activation level of -1 (Wilson, 2012).

Epoch is a term used to describe a complete pass through all of the training patterns (one cycle). The weights in the neural net may updated after each pattern presented to the net, or they may updated just once at the end of the epoch (Wilson, 2012).

Hidden layer is a layer of neurons connected between input layer and output layer. Usually structured into two or more layers. The neurons in each layers is connected to all neuron in the following layer (Wilson, 2012).

Hidden unit is a hidden unit in a neural network is a neuron which is neither an input unit nor an output unit (An-Najjar, 2005).

A learning algorithm is a systematic procedure for adjusting the weights in the network to achieve a desired input/output relationship, i.e. supervised learning (Wilson, 2012).

5.3 Mechanism of ANNs
Generally, the ANN trained so that each input data lead to a specific output. Then the network adjust the weight of the factor depending on comparing process between the real output and the calculated one (An-Najjar, 2005).

![Figure (5.2): The Concept of ANN (An-Najjar, 2005)](image)

5.4 Types of ANNs
There are three major types (An-Najjar, 2005):
- Single-Layer Feed forward Networks: It is the simplest type of ANN, it does not have any hidden layer.
- Multilayer Feed forward Networks: it contains one or more hidden layer.
Recurrent Networks: It consists of a closed loop, where the output layer feedback the input layer. It is developed to deal with the time varying or time-lagged patterns

5.5 ANNs in traffic safety analysis

Kunt et al. (2011) make a prediction model for traffic accidents severity using ANN with Genetic Algorithm (GA). He used a huge number of factors just as driver gender and age, using of seat belt, type and safety of vehicle and traffic flow. The ANN used multi-layer perceptron (MLP) architecture that consisted of a multi-layer feed-forward network with hidden sigmoid and linear output neurons that could also fit multi-dimensional mapping problems arbitrarily well. The network has 12 input, 25 neurons in the hidden layers and 3 neurons in the output layer. They claimed that the advantage of these models is improving themselves adding new data.

Akgüngör (2008) used ANN to make a prediction model using data from 1986 to 2000 he used AADT, fatalities, injuries, accidents and population.

Al-Ghirbal (2005) used an ANN model to predict traffic accident rate at roundabout in using ANN. He built the model by relating the available geometric (angles between the approaches, size of the central island, flaring entry, the number of roundabout’s approaches, etc…), traffic characteristics (entering and circulating traffic volumes, etc…) and accident records. The sensitivity analysis proved the capability of the model to recognize and generalize complex nonlinear relationships. And he recommends utilizing this approach to model other traffic problems.

Chong et al. (2004) model the severity of traffic accidents using ANN with decision tree. He used the following factors: drivers’ age, gender, alcohol usage, restraint system, eject, vehicle body type, vehicle age, vehicle role, initial point of impact, manner of collision, rollover, roadway surface condition, light condition, travel speed and speed limit. The model gives 75% prediction of fatal accidents.

Benedetto (2004) used ANN to improve road design and rehabilitation by prediction traffic accidents. He used the AADT, environment factors and human factors in prediction models.

Thurston (2002) used the ANN with GIS to build a thinkable GIS using artificial intelligent and use traffic flow. This may increase the capabilities of ANN to make spatial analysis. He claims that using ANN with GIS will make all variables in traffic flow with respect to the any accident could be processed resulting in a determination of optimum re-routing until the traffic flow is stabilized. Therefore, through the application of an ANN, GIS professionals can add another dimension to their spatial capabilities.
6 TRAFFIC ACCIDENTS IN PALESTINE

This chapter contains a general view of traffic accidents in Palestine, especially in Gaza, in the period from 2000 to 2013.

6.1 Data Resources

The accidents data used in this research are obtained from the Ministry of Transportation (MOT), Gaza Police Department (GPD), Palestinian Central Bureau of Statistics (PCBS) and the land authority.

Data obtained from MOT includes accidents that have been turned to the courts in the period from 2000 to 2005. The data includes the following:

1. Accident data: which include the exact site location, date and the conditions that causes the accident.
2. Injuries data: which includes injury type for persons included in the accidents and connected to the accident that caused it.
3. Driver data: which includes the driver’s name, ID and license number.
4. Vehicle data: which includes the type of vehicle and the insurance company.

The research does not include data after 2005 because the computerized archive was stopped in 2006. This was due to the political situation and the Political division in Palestine.

The researcher tried to include more data from the Public Prosecution, but there were some obstacles that prevented this step which can be summarized as follow:

1. The general situation in Gaza Strip and the repeated attacks by occupation forces on the security services
2. The huge number of files.
3. The Public Prosecution prevented the researcher to transfer files to a safe place to computerize them because the data is classified as personal and private data.
4. The time limit of the thesis.

The road data and maps used in the research are also obtained from MOT. It contains main, secondary and residential roads all over Gaza strip. It also contains the speed limits the streets, the surface of the street, number of lanes, the width of the street, the median existing and the type of control on the street if it is one way street or two way street.

The aerial photo for Gaza Strip used in the GIS program is obtained from the Land Authority. It is 0.5 m x 0.5m Pixel and it is taken in 2008.

The accident statistics in the West Bank from 2001 to 2013 and Gaza Strip from 2001 to 2007 are obtained from the PCBS publications. While the accident statistics in the Gaza Strip from 2007 to 2013 are obtained from the GPD.

6.2 Traffic accidents in Palestine

It is easy to note that in the last few years, the number of traffic accidents and the number of accident casualties are increasing rapidly in Palestine, especially in Gaza Strip, where both of them increased many times from 2007 to 2013.
Tables (6.1) and (6.2) show that number of registered traffic accidents oscillates up and down in the period from 2001 to 2007; this may refer to the political situation in Palestine “Intifada” where occupation forces divided the Palestinian areas to isolated areas and controlled the traffic movements between them. In 2007 and 2008, the number of accidents in Gaza Strip has dropped to the half while the number of the fatality did not change, this may refer to the new political situation (Palestinian division) which led to reconstruction process of the security agencies and create new trend to encourage the amicably resolved accidents and magistrate's clan. Noting that GPD don not archive amicably resolved, this process affected the annual number of accidents.

Table (6.1): Traffic Accidents and Casualties in Palestine.

<table>
<thead>
<tr>
<th>Year</th>
<th>Adjudge accidents</th>
<th>Casualties</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Slight</td>
</tr>
<tr>
<td>2001*</td>
<td>5015</td>
<td>3377</td>
</tr>
<tr>
<td>2002*</td>
<td>2503</td>
<td>1352</td>
</tr>
<tr>
<td>2003*</td>
<td>3978</td>
<td>2395</td>
</tr>
<tr>
<td>2004*</td>
<td>4722</td>
<td>2992</td>
</tr>
<tr>
<td>2005*</td>
<td>5628</td>
<td>4351</td>
</tr>
<tr>
<td>2006*</td>
<td>3509</td>
<td>3950</td>
</tr>
<tr>
<td>2007**</td>
<td>3776</td>
<td>3600</td>
</tr>
<tr>
<td>2008**</td>
<td>4427</td>
<td>3629</td>
</tr>
<tr>
<td>2009**</td>
<td>6211</td>
<td>6000</td>
</tr>
<tr>
<td>2010**</td>
<td>6503</td>
<td>8297</td>
</tr>
<tr>
<td>2011**</td>
<td>10950</td>
<td>9089</td>
</tr>
<tr>
<td>2012**</td>
<td>11912</td>
<td>9635</td>
</tr>
<tr>
<td>2013**</td>
<td>10466</td>
<td>8207</td>
</tr>
</tbody>
</table>

* PCBS ** GPD and PCBS

Table (6.2): Traffic Accidents and Casualties in Gaza.

<table>
<thead>
<tr>
<th>Year</th>
<th>Accidents</th>
<th>Casualties</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adjudge</td>
<td>Amicably resolved</td>
</tr>
<tr>
<td>2001*</td>
<td>1538</td>
<td>N/A</td>
</tr>
<tr>
<td>2002*</td>
<td>1504</td>
<td>N/A</td>
</tr>
<tr>
<td>2003*</td>
<td>1778</td>
<td>N/A</td>
</tr>
<tr>
<td>2004*</td>
<td>1786</td>
<td>N/A</td>
</tr>
<tr>
<td>2005*</td>
<td>1930</td>
<td>N/A</td>
</tr>
<tr>
<td>2006*</td>
<td>1329</td>
<td>N/A</td>
</tr>
<tr>
<td>2007**</td>
<td>842</td>
<td>N/A</td>
</tr>
<tr>
<td>2008**</td>
<td>752</td>
<td>N/A</td>
</tr>
<tr>
<td>2009**</td>
<td>1540</td>
<td>2277</td>
</tr>
<tr>
<td>2010**</td>
<td>1095</td>
<td>4171</td>
</tr>
<tr>
<td>2011**</td>
<td>3703</td>
<td>4608</td>
</tr>
<tr>
<td>2012**</td>
<td>3955</td>
<td>5897</td>
</tr>
<tr>
<td>2013**</td>
<td>2639</td>
<td>2739</td>
</tr>
</tbody>
</table>

* PCBS ** GPD
The tables also show that the number of casualties in Palestine also oscillates up and down in the period from 2001 to 2008, while it increased more than 60% in the years after.

In Gaza, the increment in fatal accidents may refer to the huge number of motor cycles (more than 10,000) that are merged to the traffic movement in Gaza without proper training programs to qualify drivers.

It is important to illustrate that, in Palestine, the number of slight casualties has the largest increment rate, which increases more than 240% from 2001 to 2013, and followed by fatal casualties with 44% increment in the same period.

![Figure (6.1): Adjudge accidents vs amicably resolved accidents in Gaza](image)

It is very important to illustrate that the number of amicably resolved accidents in Gaza is more generally more than adjudge accidents. Figure (6.1) shows that the number of amicably resolved accidents may reach four times the adjudge accidents. This situation is a normal result to the police policy to include the magistrate's clan as a part of traffic accident causes.

According to GPD (2013) and PCBS (2012), the number of registered vehicles in Palestine in 2011 is 194,738 vehicles. This means that there are 21 inhabitants for each vehicle in Palestine. With 5.8 inhabitants as an average family size, each 3.6 families owned a vehicle. These numbers reflect the economic and social situation in Palestine.

Figures (6.2) and (6.3) show the causality rate (CR), causality per accident, in Palestine and Gaza Strip respectively. In general, CR has a stable behavior from 2001 to 2013. However, from 2006 to 2010. However, in 2009, the CR value in Gaza Strip dropped to 0.48. This may refer to non-recording slight injuries due to political situation.

In 2010, the CR in Gaza increased to 2.1 causality/accident that is twice the average. This may refer to the motorcycles accidents that increase the risk of death in traffic accidents in Gaza.

As shown in Figure (6.4) and Figure (6.5), the value of Fatal Causality Rate (FCR), fatal causality /100 accidents, in Gaza is higher than the total Palestine values. According to Table (6.1) and Table (6.2), the casualties in Gaza represent from 9-45% of total casualties in Palestine (with average of 30%), while it represent from 34-73% of fatal accidents (with average of 46%) and the accidents in Gaza represent from 16-60%
of total accidents in Palestine (with average of 32%). These values reflect that the accidents in Gaza have risk factors higher than average in Palestine.

Figure (6.2): Causality Rate (CR) in Palestine

![Causality Rate in Palestine](image)

Figure (6.3): Causality Rate (CR) in Gaza

![Causality Rate in Gaza](image)

Figure (6.4): Fatal Causality Rate (FCR) in Palestine

![Fatal Causality Rate in Palestine](image)
Figure (6.6) shows that in 2011, the fatality risk per 100,000 populations in Palestine is five. This number is very small compared to other Arab and regional countries. This situation may be explained by the huge number of population per vehicle in Palestine and that more than 50% of Palestinian people are children (under legal driving age).

![Figure (6.5): Fatal Causality Rate (FCR) in Gaza](image)

On the other hand, Figure (6.7) shows that the number of fatalities per 10,000 vehicles in Palestine is 10.7. This is a very high rate compared to other countries. This rate reflects the increase of risk in Palestine road network.

![Figure (6.6): Fatality Risk per 100,000 Inhabitants in 2011](image)

![Figure (6.7): Fatality Risk per 10,000 Vehicles in 2011](image)
6.3 Discussion of the Accidents during 2000 to 2005 in the Gaza Strip

The accidents database contains the accidents that went to the court. As mentioned by GPD, the blind accident consists of a huge number of accidents that was solved consensually.

Figure (6.8) shows that the concentrations of accidents are in the downtowns which have a high traffic volume and on the inter-city roads which include roads with high volume and velocity. This observation indicate that the flow and the velocity may be considered as main factors in traffic accidents occurrence.

Table (6.3) shows the percentage of each type of accidents in the Gaza Strip from 2000 to 2005. It indicates that more than 80% of registered accidents are PDO and slight accidents.

Figure (6.8): Accidents in Gaza Strip

Table (6.3) shows the percentage of each type of accidents in the Gaza Strip from 2000 to 2005. It indicates that more than 80% of registered accidents are PDO and slight accidents.
Table (6.4) shows that about 80% of accidents in (2000-2005) period are daylight accidents (from 06:00 AM to 06:00 PM).

Table (6.5) shows the number of accidents that involved carts, motorcycles and bicycles. They participated weakly. However, after 2007, the motorcycles accidents increased rapidly because of the huge number of imported motorcycle from Egypt after the Israeli withdrawal from the Gaza Strip. The Palestinian government restricted the import process after 2011 according to this situation.


<table>
<thead>
<tr>
<th>Accident Type</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatal Accidents</td>
<td>498</td>
<td>3.4%</td>
</tr>
<tr>
<td>Serious Accident</td>
<td>684</td>
<td>4.6%</td>
</tr>
<tr>
<td>Moderate Recognized</td>
<td>6507</td>
<td>44.1%</td>
</tr>
<tr>
<td>Slight Accidents</td>
<td>5933</td>
<td>40.2%</td>
</tr>
<tr>
<td>Property Damage Only (PDO)</td>
<td>960</td>
<td>6.6%</td>
</tr>
<tr>
<td>Not Recognized</td>
<td>159</td>
<td>1.1%</td>
</tr>
</tbody>
</table>

Table (6.4): Accidents during the day (2000-2005)

<table>
<thead>
<tr>
<th>Accident Type</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daylight Accidents</td>
<td>9063</td>
<td>79.1%</td>
</tr>
<tr>
<td>Night Accident</td>
<td>2232</td>
<td>19.5%</td>
</tr>
<tr>
<td>Not Known</td>
<td>158</td>
<td>1.4%</td>
</tr>
</tbody>
</table>

Table (6.5): Accidents involved carts, motorcycles and bicycles (2000-2005)

<table>
<thead>
<tr>
<th>Vehicle</th>
<th>Number of Accidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carts</td>
<td>103</td>
</tr>
<tr>
<td>Motorcycles</td>
<td>142</td>
</tr>
<tr>
<td>Bicycles</td>
<td>268</td>
</tr>
</tbody>
</table>

Table (6.6): Vehicles classification due to Insurance status (2000-2005)

<table>
<thead>
<tr>
<th>Number of Accidents</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insured Vehicles</td>
<td>9532</td>
</tr>
<tr>
<td>Non-insured Vehicles</td>
<td>4912</td>
</tr>
<tr>
<td>Total</td>
<td>14444</td>
</tr>
</tbody>
</table>

Table (6.6) shows that one third of vehicles included in accidents did not have insurance documents. This may be explained by the tough economic situation during the second Intifada in Gaza Strip.
Figure (6.9) shows that the distribution of accidents around the year shows a peak in summer time. This may explained due to the effect of the seaside visitors for recreation in the summer season, where the number of pedestrians and vehicles.

![Graph showing monthly accidents distribution]

Figure (6.9): The percentage of Monthly Accidents in Gaza Strip (2000-2005)

Table (6.7) shows that children less than 18 years represent more than 50% of the injuries in Gaza, this observation reflect the need of integrated traffic awareness program for children and other stakeholders.

<table>
<thead>
<tr>
<th>Age</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;6</td>
<td>3258</td>
<td>22.10</td>
</tr>
<tr>
<td>From 7-18</td>
<td>4180</td>
<td>28.36</td>
</tr>
<tr>
<td>From 19 To 60</td>
<td>5506</td>
<td>37.35</td>
</tr>
<tr>
<td>&gt;60</td>
<td>360</td>
<td>2.44</td>
</tr>
<tr>
<td>Not Recognized</td>
<td>1437</td>
<td>9.75</td>
</tr>
</tbody>
</table>

### 6.4 Databases evaluation

To improve the quality of the engineering studies, they should be built on realistic data. The major observations that can be recorded on the data and the databases can be summarized as follow:

1. Accidents Database:
   - The database depends on accidents description paragraph. Therefore, it is hard to extract data out of it.
   - The entered data by police have a huge amount of missing details. This may refer to the police interests, which are generally limited to the data that are required by the courts.
   - Some accident locations are linked to locations that are not well defined, such as a small shop. This process decreases the ability to locate the accidents later.
   - Some factors, which are considered as main factors in engineering studies, are not included in the database.
   - The accidents are not classified by casualty type.
• Accidents in different governorates may have the same accident number.

2. Roads Database:

• The database include some design variables without any mention to the actual data just like the surface type and the velocity
• Some missing data and unrealistic data should be modified.
• There is no integrated and unified system of road names among stakeholders like MOT and GPD. Therefore, it is hard to locate the data using a computerized algorithm.
7 MODEL DESIGN

In this chapter, the ANN model will be designed, verified and used to predict black spots occurrence and ranking.

7.1 Black Spot Criteria

The following criteria was adopted in Black Spot identification method;

- One-year period just like Germany, Portugal and others.
- It is observed that PDO accidents are generally not considered in black spot identification process. However, in this study, All types of accidents are considered. This step may be illustrated by the following points:
  - The poverty and the economic situation in Gaza increase the effect of PDO accidents on the people participated in the accidents.
  - Traffic control studies include PDO accidents in crash experience warrant to justify traffic signal (MUTCD, 2012)
- The use of weight function depends on the type of accidents: many studies used the weight function to identify black spot. They have different ratios for different accident types. In Portugal, serious accidents assumed to equal 100 slight accidents while in Flanders, Belgium it is assumed to equal 5 (Elvik, 2008).

In this study, the location is considered as a black spot if there are two fatal accidents, five serious casualty accidents, 10 moderate casualty accidents or 20 slight accidents or PDO accidents.

- The use of slide window with fixed length (100 m).

7.2 Attribute data preprocessing

As mentioned before, data preprocessing includes many steps. These steps can be summarized as follows:

1. Complete the missing factors:
   - Locate the traffic accident in ArcGIS: In the accidents database, the accidents specific location is included in the accident description. Therefore, a team of twenty students of the Islamic University of Gaza cooperated with the researcher to specify the accident locations in the GIS. This step was completed in 4 months period.

The team was chosen to cover all Gaza strip governorates. The distribution of the assistants is shown in Table (7.1).

Gaza governorate include about 50% of all registered accidents in the database. Therefore ten assistants from the whole team were chosen to cover it.

The database contains 11453 accident from 2000 to 2005, the details of located accidents is shown in Table (7.2).

As shown in Table (7.2), Khan Younis governorate has the maximum percentage of located accidents with 90%. It is followed by Gaza
governorate, Middle Area and Rafah respectively. While in Northern Gaza, the assistant managed to locate a very limited number of accidents, about 1.5%; the reason of that is the lack of the manpower in the Northern Gaza governorate. Therefore, the accidents in Northern Gaza governorate were excluded.

Table (7.1): Accidents in the Database

<table>
<thead>
<tr>
<th>Governorate</th>
<th>Number of assistants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northern Gaza</td>
<td>1</td>
</tr>
<tr>
<td>Western Gaza</td>
<td>5</td>
</tr>
<tr>
<td>Eastern Gaza</td>
<td>5</td>
</tr>
<tr>
<td>Middle Area</td>
<td>4</td>
</tr>
<tr>
<td>Khan Younis</td>
<td>2</td>
</tr>
<tr>
<td>Rafah</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
</tr>
</tbody>
</table>

Table (7.2): Accidents in the Database

<table>
<thead>
<tr>
<th>Governorate</th>
<th>Number of Accidents</th>
<th>Located Accidents</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northern Gaza</td>
<td>1561</td>
<td>21</td>
<td>1.35%</td>
</tr>
<tr>
<td>Gaza</td>
<td>5259</td>
<td>4367</td>
<td>83.0%</td>
</tr>
<tr>
<td>Middle Area</td>
<td>1543</td>
<td>1292</td>
<td>73.7%</td>
</tr>
<tr>
<td>Khan Younis</td>
<td>1691</td>
<td>1534</td>
<td>90.7%</td>
</tr>
<tr>
<td>Rafah</td>
<td>1119</td>
<td>561</td>
<td>50.1%</td>
</tr>
<tr>
<td>Not Recognized</td>
<td>281</td>
<td>4</td>
<td>1.4%</td>
</tr>
<tr>
<td>Total</td>
<td>11453</td>
<td>7779</td>
<td>67.9%</td>
</tr>
</tbody>
</table>

- Classifying the roads according to traffic volume: As shown in the previous studies, the traffic volume is considered as a main factor in traffic accidents occurrence. Therefore, the researcher classified the roads into three major categories as follows:
  1. Roads with low volume traffic (up to 5000 veh/day)
  2. Roads with moderate volume traffic (between 5000 and 10000 veh/day)
  3. Roads with high volume traffic (more than 5000 veh/day)

This classification was based on data and information extracted from several references including; MOT, local municipalities, Traffic Survey of Gaza (Sarraj, 2006), Ministry of Planning (MOP, 2010) and Traffic n Palestine website (PALTRANS.ORG).

2. Data cleaning: this step includes filling missing values, smoothing out noise and correcting inconsistencies in the data. These steps can be summarized as follow:
   - In traffic volume categories, the roads with moderate and high traffic volume are limited using data from the previous resources then all other roads are assumed as low traffic volume.
- More than 240 roads with unknown width. They are assumed three meter width.
- More than 108 roads with unknown number of lanes data, they are assumed to have one lane.
- There are 51 roads that have a velocity less than 20 km/hr. They are assumed to have 20 km/hr velocity.
- There are 6 roads with unknown type of surface. They are assumed to be unpaved roads.

3. Data integration: it is the process of merging data from multiple data sources.
   - Accidents classification due to casualty: Three columns are used to merge the data from injury. These columns are 1) accident number 2) governorate where the accidents took place and 3) the year when the accident happened. The injury table is divided into four folders depending on the type of casualty. All accidents are assumed to be PDO accidents, and then the accidents table is joined to the slight casualty table, all accidents that have a slight casualty is mentioned as slight accidents. This step is repeated to moderate, serious and fatal casualties respectively. The final table has the final accident classification due to casualty type. This merging process is done by ArcGIS using “Join Attribute” tool.

4. Data selection: applied to reach a reduced set of data, which is smaller than the original data, but very close and can be used to represent it.
   - The data is minimized by excluding all factors that do not have any effect on traffic accident analysis such as drivers and injuries’ names and identification numbers, insurance company details, accident description, and car numbers etc…
   - The researcher uses the data that have a location in ArcGIS and exclude other data. These data are about 68% of the whole data. So it can represent the data with reasonable quality.

5. Data transformation: routines convert the data into appropriate forms for mining; this step will be used to create black spot database instead of accidents database.

7.3 Black spot data preprocessing.

Black spot data preprocessing includes many steps is an extension to the traffic accidents preprocessing.

The first step is to integrate data using ArcGIS.

1. Using ArcGIS to integrate spatial data with attribute data:
   1. Divide the accidents data to layers depending on the accidents’ years and casualty type.
2. Make a 50 meter-buffer for each dataset and for each year find the number of accident that intersect with each dataset depending on the type of casualty.

3. Specify the roads that intersect with the spots using “Join by location” tool in ArcGIS.

4. Find the weight of each spot by multiplying each fatal accident with five, serious accident with two, moderate accident with one and other accident types with 0.5.

Table (7.3) shows the rank of the most dangerous black spot in Gaza Strip from 2000 to 2005. It also shows that five spots were located in Salah Aldeen Street.

Table (7.3): The most dangerous black spot in Gaza Strip from 2000 to 2005

<table>
<thead>
<tr>
<th>Rank</th>
<th>Site</th>
<th>Governorate</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Salah-Aldeen Meraj Intersection</td>
<td>Rafah</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>Salah-Aldeen Neae Road No. 8 intersection</td>
<td>Gaza</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>Alnusairat downtown</td>
<td>AlWosta</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>Alzawayda-Alrasheed St. intersection</td>
<td>AlWosta</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>Omar Almohktar Aljalaa intersection</td>
<td>Gaza</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>Aljalaa Alsaftawy Intersection</td>
<td>Gaza</td>
<td>18</td>
</tr>
<tr>
<td>7</td>
<td>Salah-Aldeen Alshejaeya Intrance</td>
<td>Gaza</td>
<td>17</td>
</tr>
<tr>
<td>8</td>
<td>Salah-Aldeen Almaghazy Intrance</td>
<td>AlWosta</td>
<td>16</td>
</tr>
<tr>
<td>9</td>
<td>Alazhar intersection</td>
<td>Gaza</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>Alrasheed St. Gaza Bridge</td>
<td>Gaza</td>
<td>14</td>
</tr>
<tr>
<td>10</td>
<td>Salah Aldeen Maan- UN Medical Center</td>
<td>Khan Younis</td>
<td>14</td>
</tr>
<tr>
<td>10</td>
<td>Omer Almokhtar Alqassam Intersection</td>
<td>Gaza</td>
<td>14</td>
</tr>
<tr>
<td>10</td>
<td>Omer Almokhtar near AlIsra Tower</td>
<td>Gaza</td>
<td>14</td>
</tr>
</tbody>
</table>

2. Data transformation:

1. Use a coding system to simplify the data in the ANN model as follow:
   1. Convert all accidents with weight more than 20 (4 fatal accident) to one.
   2. Divide other values by 20 and use it as a ratio.
   3. The understudying spots weight restricted from 0.25 to one. Spots less than 0.25 are considered as safe spots because it constitutes more than 80% of the data (6208 from 7754 spots) and it may cause model skewed.

2. Find the factors that affect the black spots as follow:

1. The number of intersections that intersect with each spot.

   This step is done by finding the number of roads that intersect with the spot using “Join Spatial Data tool” in ArcGIS, as follow:

Table (7.4): Number of intersection factor

<table>
<thead>
<tr>
<th>Number of Roads intersect with spot</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N &lt; 2$</td>
<td>0</td>
</tr>
<tr>
<td>$2 \leq N \leq 4$</td>
<td>1</td>
</tr>
<tr>
<td>$&gt;4$</td>
<td>2</td>
</tr>
</tbody>
</table>
2. The flow in the main road.

Table (7.5): Flow factor

<table>
<thead>
<tr>
<th>Flow Condition</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \leq 5000 )</td>
<td>1</td>
</tr>
<tr>
<td>( 5000 &lt; ADT \leq 10,000 )</td>
<td>2</td>
</tr>
<tr>
<td>( &gt;10,000 )</td>
<td>3</td>
</tr>
</tbody>
</table>

3. The existence of a median in the main road.

Table (7.6): Median factor

<table>
<thead>
<tr>
<th>Median Condition</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does not exist</td>
<td>0</td>
</tr>
<tr>
<td>Exist</td>
<td>1</td>
</tr>
</tbody>
</table>

4. The existence of one way road in the spot area

Table (7.7): Existence of one-way road factor

<table>
<thead>
<tr>
<th>Existence of one way road</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does not exist</td>
<td>0</td>
</tr>
<tr>
<td>Exist</td>
<td>1</td>
</tr>
</tbody>
</table>

5. The main road surface type.

Table (7.8): Surface factor

<table>
<thead>
<tr>
<th>Road Condition</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unpaved</td>
<td>0</td>
</tr>
<tr>
<td>Paved</td>
<td>1</td>
</tr>
</tbody>
</table>

6. The main road design speed.

7. Number of lanes on the main road.

8. The main road width.

7.4 Making ANN model using Visual Gene Developer program (VGD)

The researcher uses Artificial Neural Network (ANN) tools in Visual Gene Developer program (VGD) to estimate the accidents.

The data of the years from 2000 to 2003 (810 spots, about 70% of the data) is used in the model generation. The data of the years 2004 to 2005 (352 spots about 30% of the data) will be used to verify the model.

Many models are developed to optimize the model. The following model represents the optimum model created in this research:
The topology setting:
1. Eight Input Variables:
2. One output Variable: the weight of the spot.
3. Number of hidden layers: using two hidden layers with 10 nodes in each one.

The training setting:
1. Learning rate is 0.0001
2. Momentum coefficient is 0.1
3. Transfer Function is Hyperbolic tangent.
4. The maximum number of training cycles is 6,000,000
5. Target error is 0.001
6. The initialization of threshold and weight factors are random.
Figure (7.3): Network Map Analysis

The map shows that factors of Traffic volume (2) and number of lanes (7) are the most effective factors on the probability of black spot existence.

Table (7.9): Running model details

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Cycles</td>
<td>6,000,000</td>
</tr>
<tr>
<td>Sum of Squared Error</td>
<td>18.7</td>
</tr>
<tr>
<td>Processing Time</td>
<td>72 hour</td>
</tr>
</tbody>
</table>

7.5 Model Validation and Evaluation

As mentioned before, the data of 2004 and 2005 (352 spots) are used to verify the model. In model evaluation, the following diagnostic tests are calculated to find the performance of the model:

1. The sensitivity: it is the ratio between the number of correct positives to the number of total positives.
2. The specificity: it is the ratio between the number of correct negatives to the number of total negatives.

<table>
<thead>
<tr>
<th></th>
<th>Correct Positive</th>
<th>Correct Negative</th>
<th>False Positive</th>
<th>False Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>91%</td>
<td>79%</td>
<td>6</td>
<td>56</td>
</tr>
</tbody>
</table>

The sensitivity and specificity show that the model is high performance.

3. Using linear regression: Figure (7.4) shows the comparison of the black spot from the data analysis and those obtained from the trained neural network. These comparisons show that the predicted data using the trained ANN is in good agreement with the experimental results. Overall, it could be concluded that the trained neural networks are successful in learning the relationship between the input and output data and the factors explained 55% of the variance in black spot occurrence.
Discussion of the results

The VGD program does not give the factor matrix, therefore, and to specify the effect of the factor in the probability of black spot occurrence, the researcher used the “predict tool” to find the effect of each pair of factors. The legend of these relations shows in Figure (7.5).

The effect of each pair of factors as follows:

1. The effect of the number of intersections that intersect with the each spot and other factors:

- The flow in the main road: these factors have a high influence on the model when the number of intersections have a high value with moderate flow value.

- The existence of a median in the mail road: Small effective values.
- The existence of one-way road: the existence of one-way road has a high influence on the model without any attention to the number of intersections.

<table>
<thead>
<tr>
<th>One-way existence</th>
<th>No. intersections</th>
</tr>
</thead>
</table>

- The main road surface Type: these factors have a high influence on the model when the number of intersections has a high value and the surface type have a moderate value.

<table>
<thead>
<tr>
<th>Road surface</th>
<th>No. intersections</th>
</tr>
</thead>
</table>

- The main road design speed

| Small effective values. |

- Number of lanes on the main road: the number of lanes has a high influence on the traffic accidents without any attention to the number of intersections.

<table>
<thead>
<tr>
<th>No. of lanes</th>
<th>No. intersections</th>
</tr>
</thead>
</table>

- The main road width

| Small effective values. |

2. The effect of flow in the main road and other factors:

- The existence of a median in the main road: generally the influence of the flow increases in the middle values and increases when the factor of existing median increases.

<table>
<thead>
<tr>
<th>Median existence</th>
<th>Main road flow</th>
</tr>
</thead>
</table>

- The existence of one-way road: generally, the influence of the flow increases in the middle values and increases when the factor of existing of one-way roads increases.

<table>
<thead>
<tr>
<th>One-way existence</th>
<th>Main road flow</th>
</tr>
</thead>
</table>
- The main road surface type: generally, the effective of the flow increases in the middle values and increases when the factor of surface type increases.

- The main road design speed

- Number of lanes on the main road: the numbers of lanes has a high influence on the model without any attention to the number of intersections.

- The main road width

3. **The effect of the existence of a median and other factors:**

- The existence of one-way road: the number of lanes has a high influence on the model and it is increases due to the existence of median.

- The main road surface type: these factors have the ultimate influence when there is no median and the road is paved.

- The main road design speed

- Number of lanes on the main road: the median existence decreases the influence of the number of the lanes.

- The main road width
4. **The existence of one-way road:**

<table>
<thead>
<tr>
<th>Road surface</th>
<th>One-way existence</th>
</tr>
</thead>
<tbody>
<tr>
<td>The main road surface type: the influence of surface factor increases, as one-way road exist.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Design speed</th>
<th>One-way existence</th>
</tr>
</thead>
<tbody>
<tr>
<td>The main road design speed: the design speed generally decreases the influence of the existence of one-way roads. The main effect happened when the factor of one-way road is high and the speed is too small.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No. of lanes</th>
<th>One-way existence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of lanes on the main road: the effectiveness of number of lane factor increases as the one-way road exist.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Small effective values.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The main road width:</td>
</tr>
</tbody>
</table>

5. **The main road surface type**

<table>
<thead>
<tr>
<th>Design speed</th>
<th>Road surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>The main road design speed: the effective of surface type factor increases as the main road width increases.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No. of lanes</th>
<th>Road surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of lanes on the main road: the influence of the number of lanes factor is very high and increasing as road width increased.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Small effective values.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The main road width:</td>
</tr>
</tbody>
</table>
6. **The main road design speed:**

- Number of lanes on the main road: the main road design speed decreases the number of lanes effect.

<table>
<thead>
<tr>
<th>No. of lanes</th>
<th>Design speed</th>
</tr>
</thead>
</table>

- The main road width: Small effective values.

7. **Number of lanes on the main road:**

- The main road width: the numbers of lanes have a high influence on the traffic accidents without any attention to the main road width.

<table>
<thead>
<tr>
<th>Main road width</th>
<th>No. of lanes</th>
</tr>
</thead>
</table>

From these photos, we can conclude the following points:

1. The number of the lanes in the main road has the most influence on the model.
2. Generally, traffic flow has high effects especially moderate flow.
3. The design speed and the width of the main road have the least influence on the model. This may refer to using design values instead of actual values.
4. The existence of one-way road, median existence and surface type has a moderate influence on the model.
8 CONCLUSIONS AND RECOMMENDATIONS.

This chapter discusses the main conclusions, recommendations and new future trends of black spot analysis in Gaza Strip.

8.1 Conclusions

The traffic safety situation in Palestine is not stable. This situation generally refers to the political situations that affect the Palestinians. The barriers, closures and poverty are the major factors that affect traffic accidents.

Importing about 10,000 motorcycles, after the Israeli withdrawal from Gaza in 2005, changed the traffic composition in Gaza and increased the fatal causality accidents.

The known amicably resolved accidents are generally more than the adjudge accidents, but they are not recorded in accidents database. This situation decreases the ability to find practical solutions to accidents.

Children (less than 18 years) represent about 50% of the total casualty.

The black spot identification step shows that many black spots were located on Salah Aldeen Street, including five of the most dangerous ones. This reflects the importance of adopting a new accident reduction plan using the Route Action Approach.

The application of Artificial Neural Networks (ANN) to predict the probability of black spot occurrence has been investigated in this research. An ANN model is built, trained and tested using the available test data created by using historical accident data from 2000 to 2005.

The model has a very high performance; the performance is measured by the sensitivity (91%) and the specificity (72%). Also, the coefficient of determination ($R^2$) value is 0.55 which means that the factor used in the model represents 55% of the variance in black spot occurrence.

The traffic flow and the number of lanes have the main effect on the ANN black spot model.

8.2 Recommendations

1. Recommendation for Gaza Police Department and Ministry of Transportation:
   - It is important to adopt new policies that ensure recording all types of accidents to improve traffic safety programs.
   - Integrating governmental databases will improve the efficiency of the recording system and improve the quality of data in the database.
   - Computerized accidents recording system using modern technologies such as smart phones will improve the time and the quality of recording process, besides improving the quality of data.
   - It is important to record the exact location of traffic accidents using the new technical facilities such as Global Position System (GPS).
• It is important to organize more awareness programs that target children and their families to reduce their losses.

2. Recommendation for future studies:

This research has very promising results in the prediction of black spot occurrence. However, the following points are recommended for future studies to improve the results of this study:

• It is recommended to utilize other artificial intelligence techniques such as fuzzy logic or genetic programming.

• Creating a model with new factors that reflect actual data, instead of design data, would improve the quality of the model. Factors such as actual spot speed on the road, the length of the tire marks and the actual width.
REFERENCES

- Ackoff R., 2009, Pyramid to Wisdom, NC State University, USA.
- A. Benedetto C. Benedetto M.R. De Blasiis, 2004, A new effective approach to accidents prediction to improve roads’ design and rehabilitation, Advances in Transportation Studies, IV(A), 5-32 ARACNE
- CSO, 2011, Digest of Road Transport & Road Accident Statistics 2010, September, Mauritius.
• Giudici P., 2003, Applied Data Mining, Statistical Methods for Business and Industry, Wiley, Italy
• Han J., 2012, Data Mining Concepts and Techniques, 3rd edition, Morgan Kaufmann Publishers is an imprint of Elsevier, USA.
• Issa, M. 2008, Traffic Safety Reality and Hopes, Naif Arab University for Security Sciences (NAUSS), KSA.
• Michelin, 2013, Practical guide to road safety awareness raising and education in Europe, Road Safety for young people in Europe, European Commission.
• Mikulik, J. et al, 2007. Road accident investigation guideline for road engineers, PIARC.
• Ministry of Planning, Sectoral Plan (Transportation Sector), 2010.
• MOT, 2005, Palestinian Insurance Law, MOT, Palestine.
• Norman, L.G., 1962, Road Traffic Accidents Epidemiology, Control and prevention, WHO, Geneve.


• Sørensen, M., Elvik, R., 2007, Black Spot Management and Safety Analysis of Road Networks-Best Practice Guidelines and Implementation Steps, 6th Framework Programme RIPCORD-ISEREST-Deliverable,


• Szczuraszek, 2000, Road Safety Improvement Supported by Road Management Information System on the Example of a few Polish cities, Roads safety on three continents conference, South Africa.

• UNECE, 2009, Illustrated Glossary for Transport Statistics, UNECE.


• Vistisen D., 2002, Vistisen D. Models and methods for hot spot safety work, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Lyngby


ANNEX 1: Sample of the Data Used in the Model and Model Validation

- Input 1: The number of intersections that intersect with each spot.
- Input 2: The flow in the main road.
- Input 3: The existence of a median in the main road.
- Input 4: The existence of one way road in the spot area
- Input 5: The main road surface type.
- Input 6: The main road design speed.
- Input 7: Number of lanes on the main road.
- Input 8: The main road width.
- Output 1: the weight of the spot.

![Neural Network - Training Pattern](image-url)
ANNEX 2: Traffic Accident Statistics from GPD (2003-2014)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>171</td>
<td>132</td>
<td>138</td>
<td>271</td>
<td>112</td>
<td>43</td>
<td>89</td>
<td>410</td>
<td>3889</td>
<td>3843</td>
<td>4147</td>
<td>5522</td>
</tr>
<tr>
<td>Deaths</td>
<td>80</td>
<td>76</td>
<td>64</td>
<td>71</td>
<td>71</td>
<td>68</td>
<td>68</td>
<td>47</td>
<td>28</td>
<td>22</td>
<td>31</td>
<td>43</td>
</tr>
<tr>
<td>Injuries</td>
<td>1048</td>
<td>144</td>
<td>80</td>
<td>320</td>
<td>321</td>
<td>41</td>
<td>35</td>
<td>118</td>
<td>79</td>
<td>97</td>
<td>88</td>
<td>65</td>
</tr>
</tbody>
</table>

Note: The table above provides traffic accident statistics from the GPD for the years 2003 to 2014. It includes data on cases, deaths, and injuries.