2005

Artificial neural network in classification of severity levels in crashes with guardrail

Fouad N. Shoukry
West Virginia University

Artificial Neural Network in Classification of Severity Levels in Crashes with Guardrail

Fouad N. Shoukry

Thesis
Submitted to the College of Engineering and Mineral Resources at West Virginia University in partial fulfillment of the Requirements for Degree of Masters of Science in Civil Engineering

David R. Martinelli, Ph.D., Chair
Ronald Eck, Ph. D.,
Darrell Dean, Ph.D.

Department of Civil and Environmental Engineering
Morgantown, West Virginia
2005

Keywords: Guardrail, Neural Network, Crash Severity

ABSTRACT

Artificial Neural Network in Classification of Severity Levels in Crashes with Guardrail

Fouad N. Shoukry

This research focuses on using artificial neural networks to classify the severity levels of crashes involving guardrails, and to subsequently identify the most significant variables explaining severity in such crashes. Most of the existing research in analyzing guardrail crashes employs statistical analysis to measure severity of crashes and, unfortunately, does not incorporate much information about the factors that affect the severity concerning guardrail crashes. In the mean time, artificial neural networks have been utilized in different areas of transportation to solve engineering problems because of their ability to model non- linearity, and flexibility with large complex data sets. Data for this research were obtained from the Highway Safety Information System and were divided into two groups, the first group included roadway characteristics including guardrail/environment as input, and severity was output. The results showed that light condition, road surface condition, end and type of the guardrail significantly affect severity levels. The second group included vehicle factors and human factors as input and crash severity was output. The resulting classification was significantly affected by the driver age and vehicle impact. Merging all factors in one model resulted in the best classification of different levels of severity (above 93% in testing classification for different class of severity) and MSE = 0.027089 in cross validation. The results have demonstrated that the Neural Networks are an effective tool to classify severity levels in crashes with guardrail if appropriate input data is available.

ACKNNOWLEDGMENTS

I would like to express my gratitude to my research and academic advisor Dr. David R. Martinelli for providing the focus and direction for completion of this work His help, advice, guidance encouragement and support have been indispensable in completion of this work.

Sincere appreciation must also be expressed to Dr. Ronald Eck for his help and encouragement to complete this work. I would like to thank Dr. Darrell Dean for his time to be a member of the committee.

I would like to thank my wife for her unlimited support during the preparation of this work. I would also like to thank my brother Dr. Samir Shoukry for his encouragement.

The author gratefully acknowledges the financial support for his research from Mid Atlantic Universities Transportation Center.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

For many years, guardrails have been effectively used to shield motorists from roadside hazards. Despite their overall effectiveness, collisions with guardrails may result in significant property damage or injury, due to guardrail failure, post failure, soil settlement, or other condition resulting in the guardrail not performing its intended function. Over the years, design engineers have modified essentially every aspect of guardrails through different full-scale crash tests in an effort to improve their effectiveness and reduce the likelihood of serious property damage and injury. These standard crash testes were presented in National Cooperative Highway Research Program (NCHRP) Report Number 230 (1) which recommended two tests on standard sections of barriers: one with an 1800 lb. vehicle impacting at 60 m/h speed at a 15 degree angle to evaluate the risk, and the other test with 4500 lb. vehicle impacting at 60 m/h and at 25 degree angle to evaluate structural integrity of the barrier. These crash tests do not reflect all types of vehicles. NCHRP Report Number 350 (2) requires that testing be done with full-size pickup trucks to better accommodate the vehicle fleet on U.S. highways. In recent years, there has been a growth in sport utility vehicles and pickup trucks on U.S. highways, thus the need to increase their representation in standard guardrail tests. The report also requires testing with very light compact cars as well as heavy trucks for proper guardrail installation. The continued evolution of guardrail design will rely on an understanding of the crashes involving guardrails to determine if there is any action that could be taken to reduce their severity based on an understanding of the factors that affect the severity of crashes with guardrails.

In the 1988 edition of its Roadside Design Guide, the American Association of State Highway and Transportation Officials, AASHTO, (3) presented warrants for the installation of traffic barriers. A traffic barrier should be installed only if it reduces the severity of crash i.e. hitting the barrier will be less severe than negotiating the hazard. The selection criteria for roadside barriers depend on many factors:

1. Barriers must be structurally able to contain and redirect the design vehicle
2. Expected deflection of the barrier should not exceed available distance to deflect
3. Barriers must undergo routine maintenance and maintenance after collision
4. Simpler designs, besides costing less, are more likely to be reconstructed properly by field personnel.

The Roadside Design Guide (3) contains the variables that should be considered in designing guardrail; those variables are identified in Figure1.1 .Subsequently, the length of guardrail can be calculated using the following equation (1):

$$X = L_H + (b/a)\ (L_1) - L_2\ /(b/a + L_H)/L_R$$

The lateral offset Y, which is defined as the horizontal distance from the edge of the road to the face of the guardrail can be calculated from the following equation (3):

$$Y = L_H - L_H/L_R\ (X)$$

Although barriers are used to shield users from obstacles located along the roadway, serious problems still exist in the highway roadside due to fixed objects.

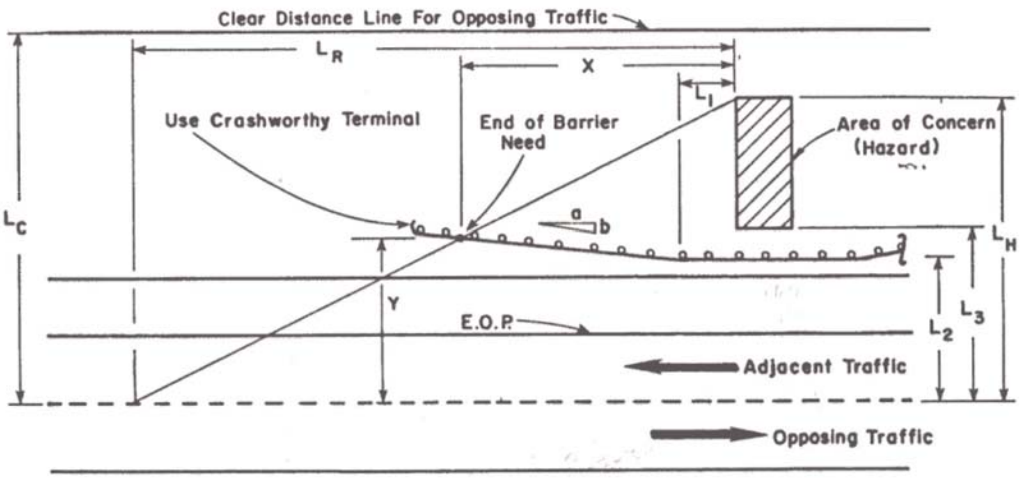

Figure 1.1 Barrier Layout Variables

X= Length of Need

Y= Lateral offset

L_R= Runout Length

L_H = Lateral Extent of the Hazard

L_C =Clear Zone

L_1= Tangent Length of Barrier

Upstream from the Hazard

L_2 = Lateral Distance of Guardrail

from the Edge of the Travel Way

Analysis of crash data indicates that more than 40% of all highway crashes involve vehicles colliding with fixed objects on the roadside. The distribution of fatal crashes with various types of fixed objects is shown in Figure 1.2.

As can be seen from the Figure 1.2, guardrails represent 12% of all fatalities involving fixed objects. Although guardrails are designed to protect motorists, the

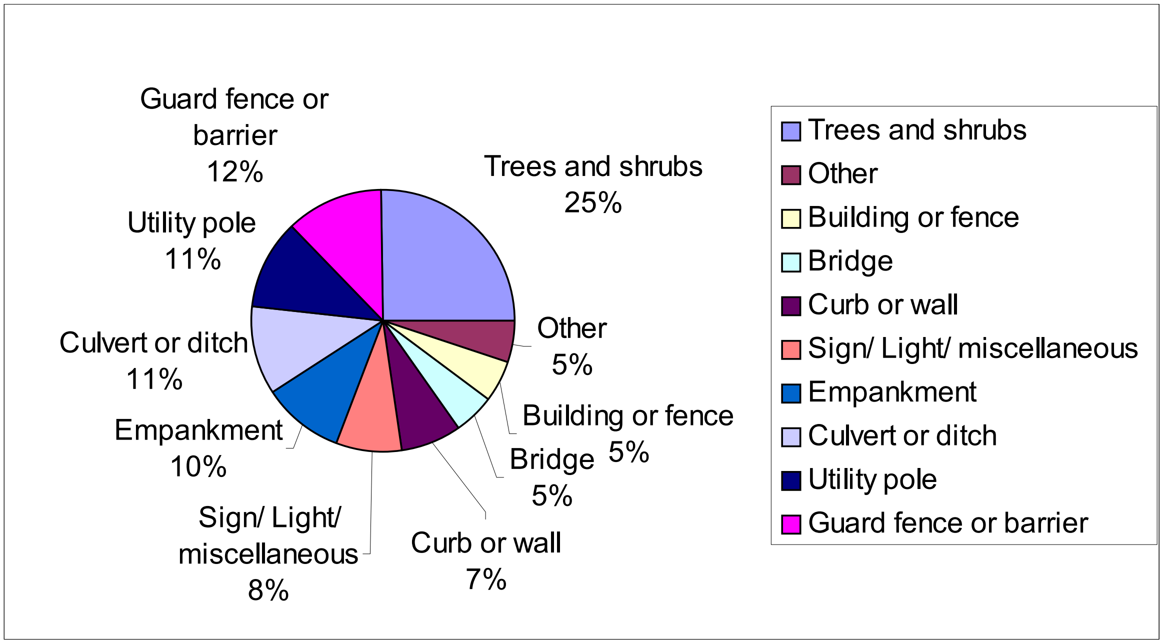

Figure 1.2 Distribution of Roadside Objects in Fatal Accidents (4).

statistics on fatalities suggest they can pose hazards. When a vehicle collides with a barrier it will result in an accident with a certain classification of severity ranging from property damage to fatality. The severity of a guardrail accident varies according to (among other things) the vehicle speed, angle of impact, and guardrail performance. This is not to suggest that guardrails are not a valuable asset to highway safety, but rather, an understanding of factors of crash severity involving guardrails could help make design and maintenance of the traffic barrier more effective.

The numerous variables and complex relationships between the driver, vehicle and roadway characteristics, guardrail design and crash severity require analytic approaches that are not conventional. This is because these factors are very different from one situation to another. For example, human factors vary from person to person due to age, gender, driver condition, reaction time and visibility. Vehicle factors are different due to the diverse nature of the vehicle fleet. The roadway contributions to the severity of traffic crashes severity varies from segment to segment due to changes in slope, design and maintenance of the roadway hardware or change in the lateral distance from the edge of the road, and other factors.

Identifying factors that map to some measure of crash severity requires special attention from the researcher in the field of transportation. Better information on the factors that affect crash severity, including the guardrail itself, could lead to modifying the design of the guardrails and as a consequence a reduction in property damage, injuries, and fatalities would result. Most of the existing work in analyzing guardrail crash data employs statistical analysis to measure the severity of crashes and, unfortunately, do not incorporate much information concerning factors related to the severity of guardrail crashes.

The limitation of available data and/or the traditional analytical techniques could be the reasons for the limited understanding of the distribution of severity levels. In a resent study, Michigan State developed and maintained a guardrail file up to 1992. In addition, the accident file and the road file can be merged with the guardrail file to generate variables that can be used in the artificial neural network model and identify the most important factors that can lead to a certain severity level in crashes with guardrails.

A novel approach to analyzing the relationship between the severity of collisions involving guardrails and the various roadway, driver, and vehicle factors, is the Artificial Neural Network. This method is unique because it gives engineers the ability to capture and represent input / output relationships to principal contributing factors that affect guardrail crashes. This thesis presents the workings of the artificial neural networks and its value to the safety of those using the highway system.

1.2 Problem Statement

Collisions with guardrails may result in significant property damage or injury. Crash tests have been conducted for many years to understand the performance of different barriers when hit by a vehicle. However, crash tests do not give an understanding of the severity level and do not consider other factors such as human factors. The object of this study is to explore the use of Artificial Neural Networks as an analysis technique for classifying different severity levels in crashes involving guardrails. Specifically, the work examines the ability of neural networks to successfully classify the severity of crashes involving guardrails with respect to human factors, roadway factors and vehicle characteristics. It is expected that neural networks, will be effective in correctly classifying the crash severity. The results could be used to provide guideline for future research on guardrail crashes.

1.3 Research Objectives:

Guardrails are a fundamental component of the roadside safety system throughout the US. Unfortunately, the circumstance by which guardrails may play a role in crash severity is not well understood. There are many possible roadway, vehicle, and driver characteristics along with guardrail factors that may, in some way, affect the safety and effectiveness of guardrails. Since guardrails are designed to protect road users from reaching hazards along the roadside it is also considered a fixed object. The shortcomings of detailed data and traditional analytical techniques could possibly be a reason for this lack of understanding. At the same time, Artificial Neural Networks have been very effective in classifying relationship between inputs and outputs, thus have emerged as an effective tool for solving problems involving complex relationships among variables.

The objective of this study is to explore the use of Artificial Neural Networks as an analysis technique for classifying different severity levels in crashes involving guardrails, and to identify the most important factors that could significantly affect their severity, specifically:

(1) To conduct a literature review on the factors which affect the severity of crashes;
(2) To analyze data obtained from the Highway Safety Information System which includes accident files, road files and guardrail files;
(3) To analyze data using Neuro Dimension software to find out the most important factors that can give the most accurate classification of severity levels in crashes with guardrail.
(4) To draw conclusions and recommendations based on the performance of the model used and its respective variables.

1.4 Thesis Organization

A review of the guardrail crashes and neural network literature is presented in Chapter 2. The sources of data collection and definition of variables are then represented in Chapter 3. Chapter 4 describes the Artificial Neural Network and analytical methodologies. Chapter 5 includes the results of the Neural Network analysis. The conclusions and recommendations from this study are presented in Chapter 6. Appendix A contains a copy of the Michigan Police crash report form. Appendix B contains results of testing and cross validation metrics produced by the neural networks that were used for calculating a weighted matrix.

Chapter 2

Literature Review

Since the study purpose is to use the neural networks to classify severity levels in crashes with guardrails, the literature review focuses on guardrail crashes, the factors related to crash severity and application of neural networks in different areas of transportation.

2.1 Guardrail Crashes

Lee Jinsun and Fred L. Mannering (5) investigated the relationships among roadway geometry, roadside geometry, roadside characteristics and run-off-roadway crash frequency and severity. The aim is to provide a basis for identifying cost -effective ways to improve highway designs that will reduce the probability of vehicles leaving the roadway and the severity of crashes. The effects of roadside features on run-off-roadway crash severity were studied with a nested Logit model. One of the roadside features that was found to significantly affect the severity of run-off-roadway crashes was guardrails. Dangerous driver behaviors such as speeding, reckless driving and driving under the influence of alcohol or fatigue can significantly increase the risk of a severe crash. The severity distribution of guardrails were 73.7% property damage, 14.0% possible injury, 10.5% evident injury, 1.8% disabling injury and 0% fatality. The study found that widening lanes, relocating roadside fixed objects, flattening side slopes and maintenance of roadside is required to reduce run-of-road crashes.

Michie and Bronstad (6) studied crashes involving guardrails. They concluded that up to 40% of guardrail crashes result in fatalities or injuries. Their conclusion was based on analysis done in New York Department of Transportation. Also, longitudinal barriers have been improperly given poor performance ratings based only on reported accident data. Using estimates of unreported accidents (90%), the success rate of longitudinal barriers is at least 94 percent, considering all types of barriers in all kinds of conditions during impacts that are within and outside the normal performance range. Guardrails are designed to perform well in crash tests. Crash test depends on three factors, which are the weight of the vehicle, speed, and the angle of approach. If the impact condition exceeds the design capacity, then barrier performance will be outside the performance range, i.e. heavy trucks, high speed, and angles clearly beyond design range.

Jovanis and Chang (7) demonstrated that conventional linear regression models are not appropriate for modeling vehicle accident events on roadways, and test statistics from these models are often problematic. This is because of the relationships between the mean and variance of accident rates, non-negativity of dependent variables, and the invalidity of normal error term distribution assumption.

Hunter, Stewart, Eccles, Huang, Council and Harkey (8) evaluated the installation of cable median barrier in North Carolina, by using historical crash data. Their analysis indicated that several kinds of crashes increased after installing cable median barrier (ran-off–road, left hit, fixed object). However, these sections showed improved overall safety due to fewer serious fatal crashes. Overall severity values were reduced after installation.

Agent (9) found that crashes with guardrails on primary and secondary highways represent two percent of all crashes. This is less than that for interstates and parkways which are only 17 percent of all crashes. This may be due to a lack of reporting of low-severity impacts in cases of primary and secondary highways, and also due to the higher exposure of freeway traffic to guardrails on interstates and parkways. However, the severity of guardrail crashes was higher on primary and secondary highways. The higher percentage of vehicles vaulting over the rail was found to be the causes for the high severity. A computer program was used to obtain the total number of accidents, injuries and fatalities which occurred on a road system during two-year period. Crashes were classified according to the number of lanes crossed before striking the guardrail to provide an estimate of the impact angle. Vehicle reaction upon impact was categorized and the severity of each category was determined by means of a severity index. The smallest possible value of the severity index is 1.00 corresponding to the case when all accidents involve only property damage, and the largest possible value is 9.5 corresponding to all accidents resulting in a fatal or injury type A. Agent also found that there is a direct proportion between the impact angle and severity. Vehicle penetration through guardrails was limited mostly to heavy trucks. During the study period, two types of end treatment were present, buried end (the end turned under the ground) and blunt (the end is exposed). The buried end treatment provides a significant improvement over the blunt guardrail end.

. Costanzo Andrea, Luingi Cicinnati and Gennaro Orsi(10) analyzed the type and severity of injuries caused to the occupants of the vehicles involved in accidental impacts against guardrails. The results of the crash tests were carried out for different types of guardrails. Concerning the trajectory and the acceleration of the vehicle during the impact as well as the damage incurred to the vehicle itself do not allow establishing any precise element regarding the possible injuries to the occupants of the vehicle. They study of the relationship between the damage to the vehicle and the personal injuries, was performed in order to find out some new solutions for improving road safety in the field of impact against the guardrails.

Gattis, Varghese, and Toothaker (11) document attributes associated with crashes in which vehicles struck the guardrail end. Their data included crashes at a variety of guardrail end types, but most ends were either exposed or turned down. They found that roughly one third of all guardrail end accidents involve an inattentive driver striking a guardrail end, the majority of the guardrail- end type crashes were property damage. For all end types combined, about one-sixth of the crashes were fatal or incapacitating injury (injury A). The severity associated with roll/vault crashes for both exposed and turned-down ends was significantly higher than severity associated with no roll/vault crashes. When a roll/vault did occur, the results were more severe with exposed ends than with turned- down ends, although the difference was not statistically significant.

Viner (12) studied the risk of rollover in ran-off-road crashes. He explored the nature and importance of vehicle rollovers in run-off-road crashes by linking accident and roadway data. The risks of rollover in run-off-road crashes are compared by land use, road type and object struck. Side slopes and ditches were found to be the main vehicle tripping mechanism involved in rollovers while guardrails/barriers was the leading fixed object cause of rollover for rural areas.

Ray, and Weir (13) evaluated in-service performance of four types of guardrails in three states (Connecticut, Iowa, and North Carolina). The types of guardrails evaluated were cable guardrail, W-beam guardrail, and strong post guardrail. Data were collected for two years and the collision performance was measured in terms of occupant injury and barrier damage. Rail height was one of the factors affecting occupant injury. They discovered that almost 90% of the collisions with guardrails were unreported. The results showed there was no statistically significant difference between the performance of the guardrail in the three states.

2.2 Factors that are related to Severity

Many studies have been conducted to investigate the relationship between vehicle crash severity and factors that are related to severity.

Renski, Khattak and Council (14) studied the impact of speed limit on crash injury severity on interstate highways in North Carolina. Single vehicle crashes were examined because they constitute a large share of injuries. The most commonly struck fixed object was found to be the face of the guardrail. The study compared crash information collected on highway segments where speed limits were increased against similar highway segments versus segments where speed limits did not increase. The paired-comparison method and the ordered model showed an increased likelihood of Class B and Class C injuries on study segments where speed limits were increased by 10 mph, resulted in a higher probability of increased severity than those increased by 5mph. No significant changes in injury severity were found for the comparison involving highway segments where speed limits were increased from 65 to 70 mph. Higher crash severity was observed when vehicles strike the face of the guardrail after speed limits were increased by 10 mph.

Council and Stewart (15) developed severity indices for various fixed objects that are struck when vehicles leave the roadway. The study suggested the need to develop a severity index based on an airbag-equipped fleet and also the need of having larger sample size.

Lassarre (16) developed time series modeling of monthly crashes totals and deaths in order to independently evaluate the effects of seat belts and speed limits on crash severity. The study found that seat belt use and average vehicle speed (of vehicles of equal mass) have only a small influence on safety.

Jones and Whitfield (17) analyzed data on accident severity with logistic regression. They explored the effects of car mass, age of the driver, and restraint use towards predicting the severity of crashes. Driver age, car mass, and restraint use were significant parameters for predicting severity. Data were collected through the New Car Assessment Program. The New Car Assessment Program was established by the National Highway Traffic Safety Administration in 1979 to provide consumers with comparisons of crash protection. During the test, an instrumented dummy wearing safety belts measure the force of impact to the chest, head, and leg. These readings are the basis of the rating. Results showed that chest deceleration was a better predictor of overall injury than head injury. Leg injury provided a significant predicting risk of injury for unrestrained occupants.

Lui, McGee, and Pollock (18) used a logistic regression approach to model the probability of fatalities conditioned on the occurrence of a crash. However, the analysis was limited to two-vehicle crashes with at least one death. The probability of a fatality was modeled as a function of driver age, driver gender, impact points, car deformation, driver safety belt status, and car mass. The results showed that males were at lower risk than females. Age had a strong positive effect on the risk. Drivers of heavy cars had a lower risk than drivers of light cars. Drivers of cars that were severely deformed in the accident were much higher risk than drivers of cars that suffered minor deformation.

Kockelman, and Kweon (19) investigated injury severity for all crashes, two-vehicle crashes and single-vehicle crashes by using data from National Automotive Sampling System. The variables chosen were severity level, vehicle type, vehicle age, driver gender, and type of crash, number of occupants in the vehicle, light condition, and day of the week. They found that a variety of factors were affecting the severity of crashes: number of vehicles involved in the crash, driver gender, vehicle type and alcohol usage. In two-vehicle crashes, they found that manner of collisions and vehicle type were important factors affecting the severity.

Kim, Nitz, Richardson, and Li (20) focused on the relationships between alcohol and crash risk. A model was built to relate the type of crash to KABCO injury scale. The KABCO scale categorizes injuries into five levels, which are:

K = Fatality

A = Incapacitating Injury

B = Non- Incapacitating Injury

C = Possible Injury

O = No Injury

They found that crash type was a significant factor in determining the degree of injury. Also the effect of seat belt use was investigated and they found that seat belt usage has an effect on the level of injury.

A study done by McGinnis, Davis, and Hathaway (21) in longitudinal analysis of fatal Run- off -Road (ROR) crashes examined how driver characteristics such as gender, age, and alcohol relate to ROR crashes. Young drivers, male drivers, drivers over 70, utility vehicles, rollover and alcohol pose special challenge for roadside safety improvement efforts. The study showed that male drivers have higher ROR crash rates than females, even after adjusting for driving exposure. Males ages 20 to 24 have ROR crash rates 3.3 times more than females of the same age. This study showed that ROR rates for teenage males are 20 times higher than for teenage females. For drivers 70 and older, these ratios are 4.5 times higher for males and 4.0 for females. Alcohol involvement in ROR crashes is nearly 50% for male drivers ages 20 to 39 and is over 50% for all drivers during dark conditions. Not all factors affecting ROR crashes are included in the study, e.g. roadway geometry, traffic volume, and guardrail presence.

A study conducted by Shankar, Mannering and Barfield (22) employed a statistical model to examine numerous variables including roadway geometry, weather conditions, and driver characteristics. The set of data that includes information on the primary cause of the crash, time of the day, location of the crash (on or off roadway, curve or straight section or grade); roadway illumination, types of roadside objects involved in the collision, and crash type. Also, the data includes weather conditions (rainy, snowy, foggy) and roadway geometry (radii and length of horizontal curves). The data also includes pavement surface condition and vehicle data (type of vehicle, restraint system used by driver and occupants), and number of occupants. Information about the driver's age and gender was also included. Their results provided evidence on the effects of environmental condition, highway design, crash type, driver characteristics and vehicle attributes on crash severity. Four types of severity levels were used in this study: property damage, possible injury, evident injury and disabling injury.

Mercer (23) investigated the effects of human factors on crash severity. These factors were alcohol usage, restraint device, driver age and gender. The study found that driver age was the most significant factor. Younger drivers were at greater risk of being involved in casualty crashes than older drivers.

Shibata and Fukuda (24) evaluated the effect of driving without a license, alcohol use, motor vehicle speed, seat belt use in the case of motor cars and helmet use in the case of motorcycles. Severity was divided into four categories, which are death, severe injury, slight injury, and uninjured. The results showed that unlicensed drivers seemed to have a higher risk in motor vehicle crash fatality, alcohol use was suggested to be a risk factor of fatality. Speed effect was significant and the effect was more critical for motorcycles. Seat belt and helmet use were found to prevent motor vehicle traffic crash fatalities among motorcar drivers and motorcyclists.

Viner (25) used economic measures to examine the objects struck in motor vehicle crashes, which are believed to cause injury or property damage. Comprehensive costs were used to combine data on fatalities, injuries and property damage. The study found that the use of comprehensive costs could reduce distortions that may occur in analysis.

2.3 Artificial Neural Networks in Transportation (ANN$_s$)

A neural network is a data-modeling tool able to capture and represent complex input/output relationships in manner similar to the human brain. Typical networks have three components input layer, an output layer, and a hidden layer. The input layer contains the data which the network must classify (or independent variables), an output layer contains the desired output (or dependent variables), and between these two layers are one or more hidden layers which do the processing. Each layer consists of neurons connected to every other neuron in the previous layer by a link that is representative of weight. Artificial neural networks have been demonstrated to be successful in solving engineering problems in the areas of classification, prediction, and function approximation. Subsequently, there are many transportation research problems that can potentially be solved with ANNs. For example, McFadden, Yang and Durrans (26) used an ANNs to predict speed on two-lane-rural highways. Data was collected for this study using a radar gun at the mid point of horizontal curves; these data were used to train the neural network model. The input consisted of geometric parameters (degree of curvature, length of curve, superelevation), and traffic parameters (lane width, total pavement width, annual average daily traffic). Their study found that ANNs offer predictive power superior to those of regression models.

Abdelwahab and Aty (27) developed an ANN model to predict driver injury severity in traffic accidents at signalized intersections. One year of crash data was used in this study and the analyses focused on two-vehicle crashes. The input data were driver age, driver gender, alcohol use, seat belt use, vehicle type (passenger car, van, pickup), speed, point of impact, day of the crash (weekday, weekend) area type (rural, urban), time of the crash(off peak, peak), light condition (daylight, dusk, street light) and weather condition. The output was severity which the researchers classified it into three levels; no injury, minor injury, and severe injury. Results of the model showed that ANNs are a promising analytical tool in predicting the severity of crashes at signalized intersections.

Sadek, and Mark (28) used ANNs for solving the inverse transportation planning problem. Their model was designed to predict zonal trip ends given the traffic volumes on the links of transportation network.

Yang, and Fengxiang (29) used ANNs pattern recognition to classify highway traffic states into some distinctive cluster centers.

Pant and Balakrishhan (30) developed a combination ANNs and a binary logit model to predict accepted or rejected gaps at rural low-volume two-way-stop controlled intersections. The results showed that ANNs can correctly predict a higher percentage of accepted or rejected gaps. ANNs has been also introduced to incident detection. Ishak and Ishak and Haitham (31) used Multi- layer and fuzzy system using real world data. These data were collected by traffic surveillance. ANN showed success over the traditional algorithms.

Teng. H, Martinelli.D.R, and Taggart.B.T (32) applied Neural Networks to traffic prediction incident detection, the results showed a good prediction model can improve the performance of incident detection. Further, ANNs have superior capabilities in emulating nonlinear systems.

2.4 Remarks:

Although there exists work that has proven reasonably successful in using traditional numerical and statistical techniques in assessing the severity of crashes, there are some studies that used human factors, others used roadway factors and some used vehicle factors. However, there was no study including all the factors that effect severity including the use of guardrails.

Very little attention has been paid to the relationships between run-off-road crashes and roadside features. Recent national statistics indicate that about one-third of fatal crashes are associated with vehicles running off the road (11). These statistics on run -off -roadway vehicular crashes indicate the continued need for research to reduce run-off –roadway crashes. Although run-off-road crashes can presented in many ways, one of the most effective forms is the use of guardrails. The data on guardrails is available for analytical study, at the same time, a promising methodology in artificial neural networks is available, presenting an opportunity to better understand the factors that most significantly affect severity in guardrail crashes.

Chapter 3

Data Collection

This chapter presents a discussion of the Highway Safety Information System (HSIS) database and the Michigan State files that were utilized in analyzing guardrail crashes. The procedure used in preparing the crash database for analysis is also discussed.

3.1 What is HSIS?

Highway engineers are continually faced with decisions concerning the design and operation of highway systems. One of the important components of the decision-making process is the potential impact on the safety of highway users. The Federal Highway Administration developed HSIS in 1987 to help understand how highway safety is affected by the design of the roadway, selection and placement of roadside hardware, the use of traffic control measures, the size and performance capabilities of vehicles, and the needs and abilities of roadway users. This understanding can be developed through an analysis of information about crashes, roadway, traffic control devices and location of hardware and obstacles on the roadside.

The HSIS is a roadway-based system that provides data on numerous police-reported crashes, roadway inventory, and traffic variables. The criterion for inclusion of a crash varies from state to state. The HSIS uses data routinely collected by states as part of the highway management system. Data is acquired annually from a select group of states, processed into a common computer format, documented, and prepared for analysis. At present, there are nine states participating in HSIS: California, Illinois, Maine, Michigan, Minnesota, North Carolina, Ohio, Utah, and Washington.

3.2 Selecting a State

The primary criteria used in selecting a state for HSIS data is the availability of a substantive inclusion of pertinent variables in reliable quantities. Choosing a state from the HSIS data files for safety analysis depends on the problem being studied and the availability of variables (in the state) that are essential for analysis. The present study requires the guardrail file to be merged with the crash file to address the severity of guardrail- related crashes.

Exploration into the availability of variables in the state database found that Michigan is the only state that had developed and maintained a separate guardrail inventory file for 1987, 1989, and 1992. Michigan Department of Transportation district offices maintained these files, and maintenance effort was stopped or greatly reduced after 1992. The guardrail file includes only guardrail factors and does not contain information on non-metal barriers such as concrete median barriers. Each record contains variables as guardrail length to the nearest foot, guardrail type (e.g. w-beam, cable barrier, etc), rail height, lateral offset from roadway edge, terminal type and flare, number of posts and post type.

3.3 Michigan Database

The Michigan database consists of the following files: Accident, Roadway Segment, Guardrail Inventory, Cross-section, Intersection, Interchange Inventory, and Traffic Control Device Inventory. The following files were used in this study: Accident, Roadway Segment, and Guardrail Inventory. The files are discussed later in this chapter. Data from the HSIS are in separate files. In order to have all information about the variables, files have to be merged together with one or more variables common in these files. For example, the accident file and vehicle file have the same case number which should be matched to indicate the same crash. The first task was merging the data files; (data was represented from HSIS in five files). Table 3.1 represents the linking variable and merging instruction.

Table 3.1 Merging Instructions for the Data Set.

FILES	LINKING VARIABLE	MERGING INSTRUCTION
ACCIDENT	Case number, Mile post, Control section	
VEHICLE	Case number, Vehicle number	Merge on Case number with accident file.
OCCUPANT	Case number, Vehicle number	Merge on Control section with Accident file, case number and vehicle number with vehicle file.
ROADLOG	Beginning milepost, Ending milepost, Control Section	Merge on control section with Accident file along with the following constraint, milepost of Accident file should be between beginning and ending milepost of Roadlog file.
GUARDRAIL	Beginning milepost, Ending milepost, Control Section	Merge on control section with Accident file satisfying the following constraint, milepost of accident file should be between beginning and ending milepost of Guardrail file.

3.4 Description of Files

3.4.1 Accident Data Files

This file consists of data collected by various police departments (city or village police and/or township police, state police, county Sheriff's department) across the state on standard statewide accident report forms and coded by the Michigan state police. (A copy of police crash reporting form is presented in appendix A. The accident file consists of three separate subfiles, the Accident, Vehicle, and Occupant. The accident subfile contains basic information on accident type, location, and types of injuries and environment. Information on each of the vehicles involved in the crash and the objects hit in the crash are presented in the vehicle subfile. The occupant subfile contains information on each injured occupant in each vehicle that was involved in the accident. Tables 3.2, 3.3, and 3.4 contain definitions of variables found in accident file, vehicle file and occupant file.

One year of accident data (1994) was considered for the present study. This was because the guardrail file was not available after 1992. The study assumed that there was no significant change in the guardrails between 1992 and 1994

Table 3.2 Definitions of Variables in Accident File

Variable	Definition
Accident Type	Type of crash: overturn, hit train, hit parked vehicle, rear- end, head- on, sideswipe
Accident investigated by	The agency investigate the crash (state police, county sheriff, township police or city village police)
Year + Case number	This is used for merging accident , vehicle and occupant files
Control Section	This is used for merging variables and it is a code for the portion of the trunk-line system where the crash occurred
County Number	Each county has number where the crash occurred
Day of month	The day the crash happened
District + Control section number	Used in linkage data with roadlog file
Drinking in accident	Whether the driver had alcohol or not
Hour of occurrence	The hour in which the crash occurred
Highway area code	codes to identify whether the crash occurred (within 150 ft north or east of the intersection, within 150 ft south or west of the intersection, driveway related, at grade crossing, median crossing related, unknown)
Highway area code and type	Same as highway area code with more details
Highway area type	Interchange area (within ramp limits in all directions), intersection area (with 150 feet in any direction from the intersection), non-intersection/interchange area. This variable changed in 1993 based on Michigan staff inputs, it appears more accurate in earlier years.
Light condition	lighting condition at the time of the crash(day light, dawn or dusk, darkness- street light, darkness- no street light)
Month of the accident	the month of the year in which the crash happened
Number of lanes	Number of traffic lanes at the crash site
Road surface condition	Condition of road surface when the crash occurred (dry, wet, snowy, icy, muddy, deberis)
Roadway classification	Classification of roadway
Severity of Accident	Crash severity definition(Fatal, Incapacitating injury, non-incapacitating injury, possible injury, property damage)
Speed limit in the crash site	The speed limit in the crash site
Speed limit posted	Whether the speed limit was posted or not

Table 3.3 Definition of Variables in Vehicle File

Variable	Description
Vehicle object hit (object 1)	First object hit in the crash
Gender of driver (drv-sex)	Gender of driver (male-female)
Year + Case number	Combination of crash year and case number used in linkage of Accident, Vehicle, and Occupants files
Driver injury	Degree of injury cased to the driver
Driver drinking information	Whether driver was under influence of alcohol or not
Driver Hazardous action/ violation	Action of the driver which led to the crash
Driver action prior to crash	Driver intent or action before the crash
Vehicle harmful event number 1	First event which happened in the crash
Vehicle harmful event number 2	Second harmful event which happened in the crash
Vehicle harmful event number 3	Third harmful event happened in the crash
Vehicle harmful event number 4	Fourth event which happened in the crash
Driver age	Age of the driver
Vehicle contribution circumstance	Contribution of driver and vehicle to the crash
Vehicle impact code	Point on the vehicle suffering worst damage
Vehicle condition	Vehicle condition after the crash (This variable is no longer accurately coded after 1991)
Direction of vehicle traveling	Direction in which each vehicle involved in the crash was traveling
Vehicle type	Type of vehicle involved in the crash
Vehicle type 2	Weight of vehicle and type involved in the crash(uncoded)

Table 3.4 Definition of Variables in Occupant File

Variable	Description
Year + Case number	Combination of crash year and case number used in linkage files
Number of occupants	Number of occupants involved in the crash
Age	Age of injured occupant
Age category	The category in which driver age lie.
Occupant degree of injury	The severity level of occupant

3.4.2 The Guardrail File

The guardrail inventory file is a guardrail-based file consisting of 12,141 guardrail sections in Michigan, which are situated on either side of the road. The data is obtained from the guardrail installation forms submitted by the private construction and design firms hired by the Michigan Department of Transportation (MDOT). Initially, road construction plans are provided by MDOT to the firms. Accordingly, the contractors are required to submit forms in the event of new guardrail installation and also in case of guardrail maintenance and relocation. Along with the forms, the contractors are required to submit the plans containing sketches of the guardrails. State personnel feed this information into the database. The file contains information on guardrail face types, end treatment types, guardrail function, and roadway type. Tables 3.5 and 3.6 contain the guardrail types and end types found in the data, the definitions of variables of guardrail file are represented in Table 3.7.

Table 3.5 Guardrail Types in the Database

Code	Face Type	Kind
AA	TYPE A	W_beam, no blockout
AD	Type AD	Type A rail on both faces of posts
BB	Type B	Type B w beam with blockout
BD	Type BD	Type BD Type B rail on both faces of posts
CC	Type C	Two rail W-beam, top rail with blockout, lower rail without blockout
CD	Type CD	Type C rail on both faces of posts
CA		Cable barrier
TT		Thrie beam barrier
EE		Other
TD		Thrie beam on both faces of posts

Table 3.6 Guardrail end types in the data base

Cable GR end
Buffered cable terminal
Buffered
Curved End shoe
Turned down
Texas twist
Anchored to bridge or barrier wall
Expose ending
Transition from another control section or ramp
Intersection radius
Minnesota bull nose
Attenuator
Terminal
Terminal cable terminal
Sentre

Table 3.7 Definition of Variables in Guardrail File

Variable	Description
Begin mile point	Beginning mile point of guardrail run. Used in linkage with other files
Ending mile point	Ending mile point of guardrail run. Used with linkage with other files
Control section	Variable used in linkage to other files
Approach end type	Type of the end of guardrail
Approach end flair	Whether the approach end is flared or not
Guardrail location	Location of the guardrail. right side, left side, continuos median.
Guardrail type	Type of guardrail face
Guardrail use	Use of the guardrail with respect to the type of road (Roadway, interchange, rest area, weight station)
Lateral distance	Lateral distance of the guardrail face from the roadway
Number of posts	Number of posts in the guardrail
Post type	Guardrail post type
Rail material	Material of the rail (Galvanized, rusty, cable)
Rail height	Height of the guardrail face
Roadway type	Type of roadway. F = Freeway, N = Non freeway

3.4.3 The Roadway File

This file contains characteristics for about 10,000 miles of roadway, which covers 8% of the highway miles in Michigan. The roadway file contains accurate information in terms of quality of data, because the data is obtained from the roadway plans and is less susceptible to human error. Table 3.8 contains variables and their definitions found in road file.

Table 3.8 Definition of Variables in Road File

Variable	Description
AADT	Annual average daily traffic
Beginning mile point of the segment	Beginning mile post of the roadway segment
Ending mile point of segment	Ending milepost of the roadway segment
Control section	Variable used in linkage to the other files
Lane width	Average lane width
Shoulder/curb type left	Type of shoulder in the left side
Shoulder / curb type right	Type of shoulder in the right side
Median type	Type of median used (No median, concrete barrier guardrail, raised island with curb, rumble strip)
Million vehicle miles traveled	Million vehicle miles traveled on road segment (created variable after 1999 for HSIS
Number of lanes	Number of basic travel lanes
Roadway type (oneway)	Type of roadway(one way, two way, divided highway, freeway)
Paved shoulder width	width of the paved shoulder
Roadside development code	For rural 1 and for urban 3
Calculated segment length	Calculated segment length based on beginning and ending mile posts.
Posted speed limit	Speed limit set at the location in miles/hour
Terrain type	Type of terrain which level or rolling
Median width	Width of median measured in feet

3.5 Codes of Crash Severity in Michigan Data:

Each level of crash severity in the data set is identified by a code as follows:

1 = Fatality

2 = Incapacitating injury

3 = Non- incapacitating injury

4 = Possible injury

5 = Property damage

The Vehicle subfile, object 1, which represents the first object struck, was chosen to extract data concerning the crashes with guardrails. All crashes which contain missing data were deleted. The cases of fatalities were very small in the main set of data (8 cases), and after eliminating all the missing information, Code 1 disappeared. Also, it was found that the data set does not contain any of Code 5 which is property damage, thus limiting the classification levels to 2, 3, and 4. The variables that were chosen and the method of eliminating the data are presented in Chapter 4.

Chapter 4

Methodology

4.1 Artificial Neural Networks

Artificial Neural Networks (ANN_s) are a powerful data-modeling tool that are able to capture and represent complex input/output relationships. The motivation for the development of neural network technology stemmed from the desire to develop an artificial system that could perform intelligent tasks in a manner similar to that of the human brain. ANN_s resemble the human brain in two ways:

(1) They acquire knowledge through learning.

(2) Their knowledge is stored within interneuron connection strengths known as synaptic weights. The true power and advantage of neural networks lies in their ability to learn these relationships directly from the data being modeled.

ANN_s are adaptive, most often nonlinearly distributed systems. This pragmatic definition emphasizes the key features of the technology. ANN_s are distributed, adaptive, generally nonlinear learning machines built from many different processing elements (PEs). Each PE receives connections from other PEs_s and/or itself. The interconnectivity defines the topology (the way the PE_s are connected together in a neural network).

The signals flowing on the connections are scaled by adjustable parameters called weights. The PE_s sums all these contributions and produces an output that is a nonlinear function of the sum. The PE_s outputs become either system outputs or sent to the same or other PE_s. $ANNs_s$ build discriminant functions from its $PE_{s.}$ The placement of the

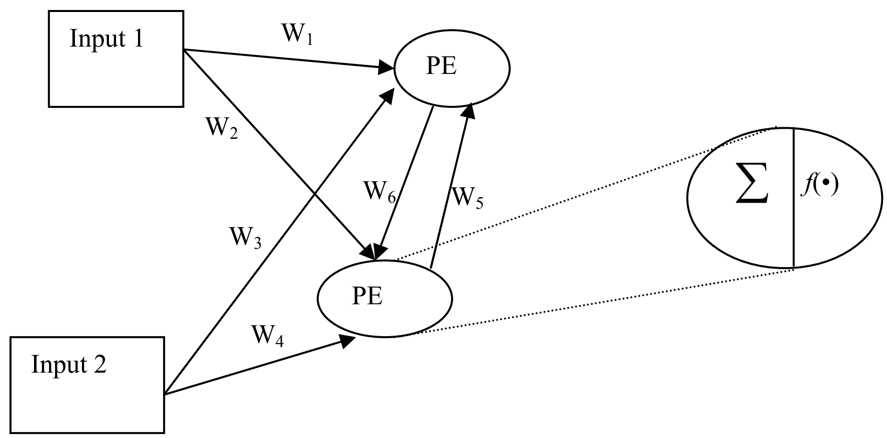

Figure 4.1 Schematic Representation of an Artificial Neural

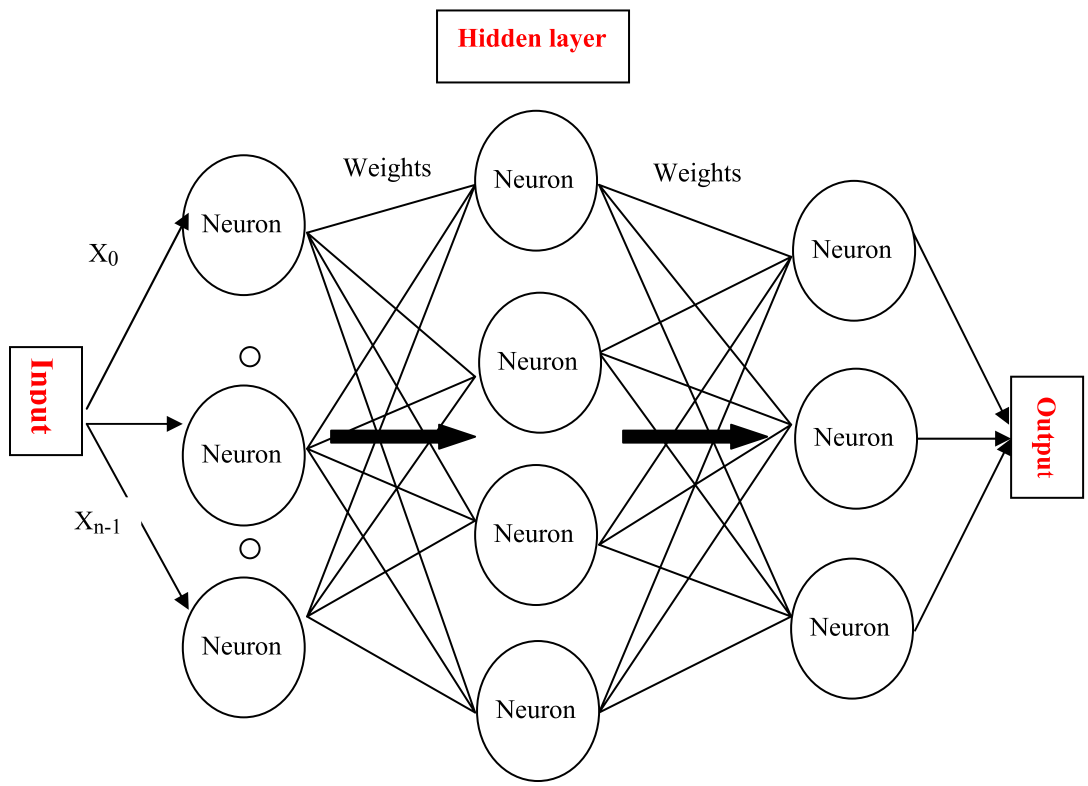

Figure 4.2 Sample of Neural Network

discriminant functions is controlled by the network weights. The weights are adjusted directly from the training data without any assumptions about statistical distribution of the data. Figure 4.1 presents a schematic of an Artificial Neural Network. Figure 4.2

presents a neural network with a structure of three layers: the input layer, the hidden layer, and output layer. For the connection between neurons on the different layers, weights are assigned which can change during training. There are many different types of neural networks, differing with respect to their architecture and transfer function. In addition, neural networks can be classified by their learning algorithm which is usually either supervised or unsupervised. A supervised network has its output compared to known answers during training. The most common type of neural network model is the multi-layer perceptron (MLP). This type of network is known as the supervised network because it requires a desired output in order to learn. The goal of this type of network is to create a model that correctly maps the input to the output using historical data so that the model can be used to produce the output when the desired output is unknown. The main advantage is that they are easy to use, and they can approximate any input/output.

The key disadvantage is that they require a relatively large data set for training. A rule of thumb (33) states that N = 10 W, that is, the training set size, N, should be 10 times larger than the number of network weights, W, to accurately classify. As mentioned before, a typical neural network can be viewed as a direct graph composed of nodes and connections (weights) between nodes. A set of vectors referred to as a training set is presented to the network one vector at a time. Each vector consists of input values and output values as shown in Figure 4.2. The inputs are X_0 through X_{n-1} and the output is Y. The goal of any neural network is to characterize a relationship between the inputs and outputs.

4.2 Backpropagation

The MLP_s learn using an algorithm called backpropagation. Basically, backpropagation is a supervised learning scheme. With backpropagation, the input data is repeatedly presented to the neural network. With each presentation, the output of the neural network is compared to the desired output and an error is computed. This error is then fed back (backpropagated) to the neural network and used to adjust the weights such that the error decreases. Thus, at every component there is a local activity and a local error so that the weights of the network can be modified. Basically, there are two equations which are:

Forward equation: $y_i = f(\sum W_{ij} Y_j) + X_i$

y_i = System output

W_{ij} = weight

X_i = Current input

Backward equation: $e_i = -\varepsilon_i + \sum W_{ij}\delta_j$

e_i = Produced error

ε_i = Injected error

δ_j= Error computes by topology

Figures 4.3 and 4.4 represent the flow calculations. In the forward pass, the network works with input data (X_i) and produces activation y_i. The backward equation works with the injected error ε_i and produces errors e_i . In more details at the ith PE the flow of activation X_i in the original network topology is from left to right, while in the topology that computes the error δ_j, it is from right to left, i.e. inputs become outputs and outputs become inputs.

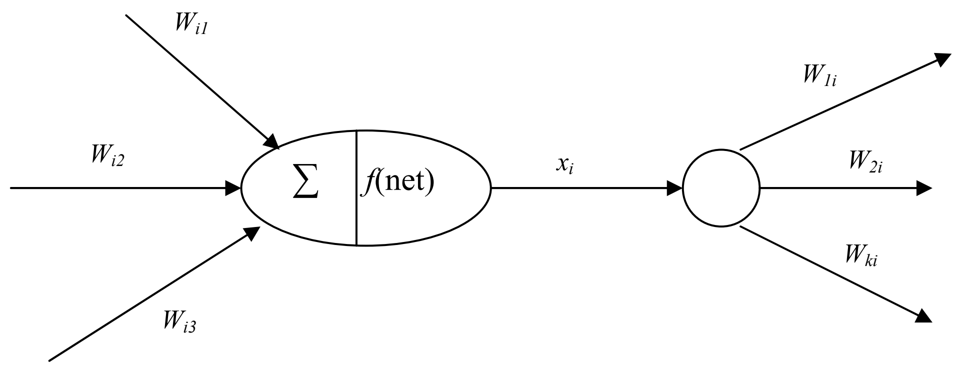

Figure 4.3 Original network

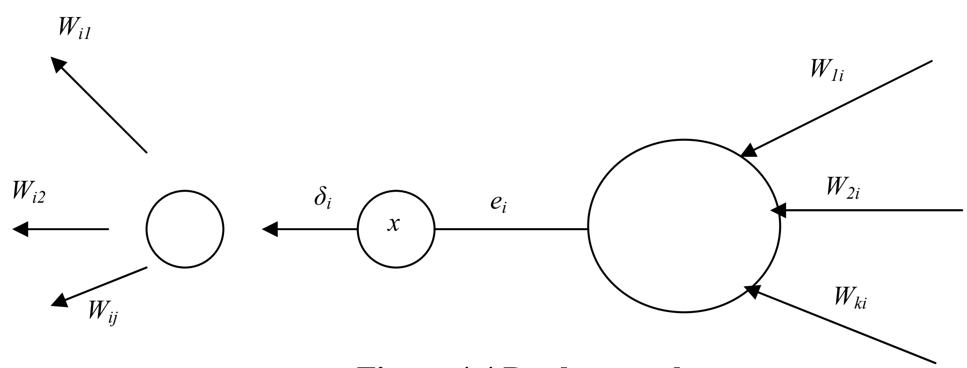

Figure 4.4 Dual network

4.3 Advantages of ANN_s for this Study

According to Principe, Euliano, and Lefebvre (34) there are six primary justifications for using Neural Networks for the analysis of this type of data set:

(1) ANNs have a remarkable ability to derive meaning from complicated data and can be used to extract patterns that are complex.

(2) Traditional linear models are simply inadequate when it comes to modeling data that contains non-linear characteristics.

(3) A trained ANN_s can be thought of as a model of function domain that can map the input to map output regardless of function form, i.e., there are no assumptions of function form necessary and a spectrum of functional forms are inherently considered (as opposed to one).

(4) An ability to learn how to do tasks based on the data given for training (adaptive learning).

(5) ANNs can be used when the data provided to solve the problem is complex and the exact mechanisms that generate the data are often unknown.

(6) ANNs are sufficiently powerful to create arbitrary discriminant functions. The ANN builds dicriminant functions from its processing elements. The ANN topology determines the number and shape of the discriminant functions. The shapes of the discriminant functions change with the topology so ANNs can achieve optimal classification. (33).

4.4 How ANN_s are used in this study

There are many types of Neural Network software. While they do not differ significantly in their performance, they may differ with respect to the visualization of the problem. One of the latest software packages is Neuro Solutions (33). This software is advantageous because it contains a demo, which helps to understand the processing of the data. It also contains icons to visualize the processing of learning, cross validation and testing procedures. Especially in the classification problem, the designer can visualize whether the data is stuck in the local minimum or not.

The Neural Expert has been used in this study. Neural Expert is a part of Neural Solution Software, which intelligently selects the network size and architecture that will most likely produce a good solution. The first step in building a neural network with the Neural Expert is to select the problem type. The classification problem has been chosen in this project since the objective is to classify crash severity. The second step is to notify the network of the location of the input file that must be in text format. The next step in constructing the neural model is to tag the input columns. The tag input columns panel is where the designer specifies which data he would like to feed into the neural network. After that, the designer has to select the desired output data. In this step, the designer can randomize the data set.

The following step is to tag the desired output columns, which for this study is severity. Then the designer shows the network which columns represent symbolic data. Finally, the designer chooses the level of generalization protection, to ensure that the network performs well on data that it has not been trained on**.** The standard method to ensure good generalization is to divide the data set into three types of data. These are:

- Training set
- Cross validation set
- Testing set

(1) Training Set is the portion of the data used to actually train the network. This is normally the largest portion of the data and can be taken as 50% of the data set. The process of learning consists of showing to the neural network examples of input data and

expected output. The synaptic weights adjust themselves until error between the output generated by network and real output reaches a desired level as mentioned before.

(2) Cross validation set is the data set aside to test the network during training.

(3) Testing set is used to further validate the results of a trained network.

4.5 Experimental design

Two steps were used for the design of the experiment

(1) Reduce the number of variables for the model from 74 variables to 13 variables based on the logic of the problem and availability of data.

(2) Reduce the number of variables for the model from 13 variables to 6 variables based on model performance under different combinations of these variables.

These two steps helped retain the variables that best explain severity of crashes by eliminating the redundant variables and the variables containing missing data. This guarantees the efficiency of the processed data, hence produces a more useful and robust model.

4.5.1 Criteria for logic-based reduction of variables

In the Vehicle subfile, OBJECT1, that represents the first object struck, was chosen to extract data concerning the crashes with guardrails. Then three steps were followed to get an uncontaminated data set:

(1) Reduce the number of variables for the model from 74 variables to 13 variables by the following procedure:

(a) Variables, which are the same, are checked from two files or more. Some of the variables that are checked are presented in Table 4.1

(b) Some variables were deleted because of missing data. Some of the variables are presented in Table 4.2

(c) Some variables were deleted because of unrelated to the problem. Examples of such variables are represented in Table 4.3

(2) Reduce the number of variables for the model from 13 variables to 6 variables based on model performance under different combinations of these variables.

Table 4.1 Examples of Variables Checked from Different Files

Variable symbol	Variable name	Files contain the variable
DWI	Driving under influence of alcohol	Accident file
INTOX	Driver drinking	Vehicle file
Drv_sex	Driver gender	Vehicle file and Occupant file
Sev	Severity	Accident file and Occupant file
Drv_ag	Driver age	Vehicle file and occupant file
Speed	Posted Speed at the crash location	Accident file and roadway file

Table 4.2 Examples of Variables Deleted Because of Missing Data

Variable symbol	Variable name	Files contain the variable
hwy_cod	Highway code	Accident file
Tot-no	Numbers of persons uninjured	Accident file
Rd-loc	Relationship to roadway	Accident file
hwy-tycd	Highway area code and type	Roadway file
vehcond 1	Vehicle condition	Vehicle file
veh-face	Vehicle contribute circumstance	vehicle file
vision	Vehicle visual obstruction	vehicle file
event1,event2,event3,event4	Harmful events	vehicle file
veh-typ	Vehicle type	vehicle file
defect	Defect of vehicle	vehicle file
lanwid	Lane width	vehicle file

Table 4.3 Examples of Variables Unrelated to the Problem

Variable symbol	Variable name	Files contain the variable
Acc-loc	Accident location	Accident file
agency	Agency investigate the accident	Accident file
Nbr-lane	Number of lanes	Accident file
spdpost	Speed limit posted (wheteter the speed posted or not)	Accident file
Dir-trv	Direction of travel	vehicle
aadt	Average daily traffic	Road file
Seg-lng	Segment length	Road file
medwid	Median wide	Road file
Sld-typ	Shoulder type	Guardrail file
Rail-mat	Rail material	Guardrail file
Grd-lngt	Guardrail length	Guardrail file

Based on this study and typical studies employing neural networks, the input variables play a critical role in achieving a good performance of classification.

Identifying the most important inputs, which have great contributions and effects on the output, is the main key to produce a good classification.

To do so, the data was divided into two groups.

The first group of data contains Roadway Characteristics/Environment, which are:

- Light condition
- Road surface condition
- Type of the end of guardrail
- Flared end of guardrail
- Guardrail type
- Post type
- Rail height
- Lateral distance from the edge of the road to the Guardrail

The second group of data contains:

- Speed limit
- Driver age
- Driving under influence of alcohol
- Vehicle impact
- Number of occupants

The Neuro-solution software was used in the experiment and a total of 53 neural networks were developed to identify the best factors that can classify the levels of severity with minimum error. As mentioned before, the data was divided into two sets. The first set was Roadway/Environment and this set contains:

Light condition, road surface condition, type of the end of guardrail, type of guardrail, flared end of guardrail, type of post, height of rail, and lateral distance from the edge of the road to the guardrail. All of these variables were chosen as input for the neural network, and the output was the severity.

The second set of data contained Human/Vehicle data and this set contains:

Driving under influence of alcohol, speed limit, driver age, vehicle impact and number of occupants.

4.5.2 Measuring Performance of Classification:

Performance of classification was measured by:

- Learning curve
- Confusion matrix for training
- Confusion matrix for cross validation

The learning curve is the plot of the mean squared error (MSE) of the network after each epoch of data. Figure 4.5 represents the shape of learning curve produced by the ANNs of the first run. The X-axis represents the epoch number and the Y-axis represents the (MSE). The MSE of the training set is shown in red, and the MSE of the cross validation is shown in blue. A network that is training well should have a constantly

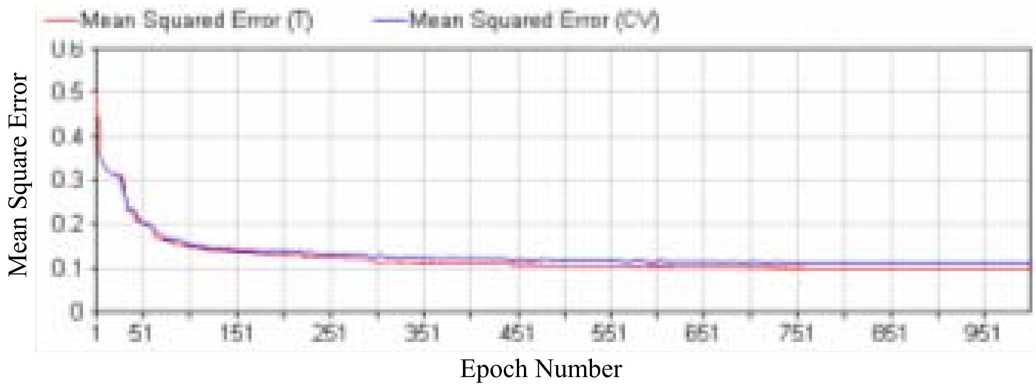

Figure 4.5 Shape of Learning Curve.

decreasing slope of the training MSE. As long as the training set learning curve is decreasing, the network is still training.

If the training set learning curve is increasing or noisy (jumping up and down), this means the network is probably not training well.

Data were examined with 672 records, but the performance of the network was very poor. The learning curve indicated that the network did not train well with the data because of lack of the quantity of data. To avoid this problem, data in an Excel spreadsheet were repeated 10 times the data set (copy, paste). Then the data became 6720 records. The neural net examined the new data and the performance of the learning curve was very good.

The confusion matrix for training and cross validation is a matrix used in classification problems to determine the number of correct and incorrect classifications by the network. The diagonal values of this matrix indicate the true classification. The off-diagonal values indicate where the neural network has been confused in classification.

Therefore, correct classifications are contained on the diagonal of the matrix [position (2,2), (3,3), (4,4)]. All other entries are incorrect classifications, but also indicate where the network was confused. Each matrix shows a percentage, but can also show absolute numbers.

The cross validation matrix is transferred into one number using the following formula:

$$Weighted\ Matrix = \frac{\%\,of\ cases(2,2)\times total\ \#\,of\ cases\ 2 + \%\,of\ cases(3,3)\times total\ \#\,of\ cases\ 3 + \%\,of\ cases(4,4)\times total\ \#\,of\ cases\ 4}{Total\ \#\,of\ cases\ 2,3\ \&\ 4}$$

The effect of eliminating each variable is discussed in chapter 5.

Chapter 5

Results

5.1 Input for Roadway/Environment Factors

The first experiment was to input the identified roadway/environment factors, which are light condition, road surface condition, end type, approach end, guardrail type, post type, rail height, and lateral distance, (8 factors) and classify the severity as output. The cross validation matrix was transferred to one number.

5.1.1 Output for Roadway/Environment

The output of the network is in the form of a matrix as mentioned before, this matrix is used in classification problems to determine the number of correct and incorrect classifications by the network. The horizontal axis indicates the true class of the input and the vertical axis indicates the network's prediction of the class. Therefore, correct classifications are contained on the diagonal of the matrix [position (2,2), (3,3), (4,4)],which are 80.4 for level 2, 83.7 for severity level 3 and 93.06 for severity level 4. All other entries are incorrect classifications, but also indicate where the network was confused. Tables 5.1, 5.2, and 5.3 represents the output of the neural network.

Table 5.1 Result of Training with Input Road/Environment in ANN

Training	2	3	4
2	80.408165	6.857143	12.734694
3	5.486399	83.725220	10.788382
4	2.110023	4.822909	93.067070

Table 5.2 Result of Cross Validation with Input Road/Environment in ANN.

Cross Validation	2	3	4
2	80.140190	2.102804	17.757010
3	3.621170	80.22839	16.155989
4	1.724138	1.724138	96.551727

Table 5.3 Result of Testing with Input Road/Environment in ANN.

Testing	2	3	4
2	75.539566	8.153478	16.3069
3	1.923077	92.170326	5.906593
4	2.755454	2.640643	94.603905

5.2 Calculation of weighted matrices

The results presented above are obtained after several runs until getting the best shape of the learning curve. To have good judgment in which factor should be eliminated, the cross validation matrix is transferred into one number using the following formula:

$$Weighted\ Matrix = \frac{\%\,of\ cases(2,2)\times total\ \#\,of\ cases\ 2 + \%\,of\ cases(3,3)\times total\ \#\,of\ cases\ 3 + \%\,of\ cases(4,4)\times total\ \#\,of\ cases\ 4}{Total\ \#\,of\ cases\ 2,3\ \&\ 4}$$

That is:

$$Weighted\ Matrix = \frac{80.14\times 1380 + 80.2\times 2410 + 96.5\times 2930}{(1380+2410+2930)} = 87.294672$$

Figure 5.1 is a copy of the first neural network to illustrate the learning curve, matrix of training and matrix of cross validation. The mean square error for this run was 0.109990

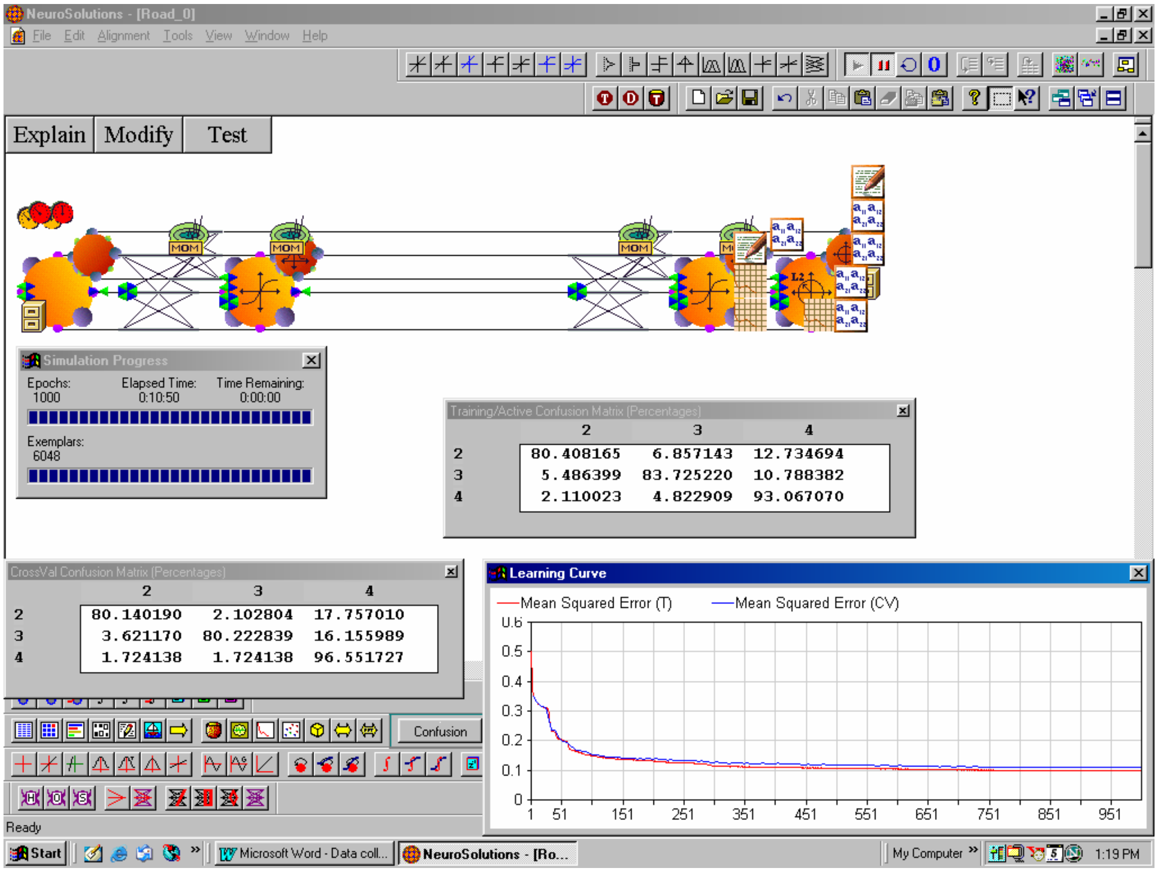

Figure 5.1 Layout of the Breadboard of ANN

5.3 Method of Modifying the Input

After this test was completed, and after taking out one variable from the data set, the network modified data was then rerun. Classification results were then obtained from the cross validation matrix. Then, the weighted matrix was calculated to indicate the effect of the deleted variable. Next, the variable that was deleted was put back and another variable was deleted. Calculation for the weighted matrix will indicate which variable has no significant effect on classification.

A total of nine neural networks were developed to conclude that the end flared variable seems to have no effect on classification. Graphical presentation of (8 variables) the results of the calculated weighted matrix are shown in Figure 5.2

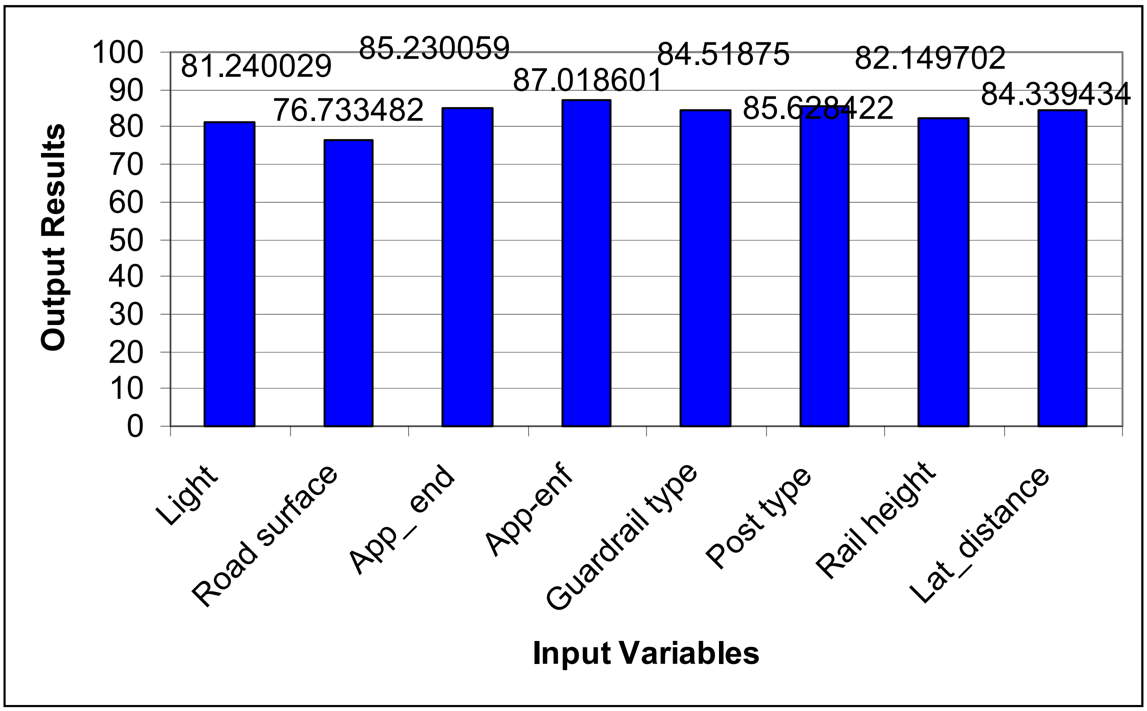

Figure 5.2 Percentage of Weighted Matrix Values for Eight Variables.

The bar chart in Figure 5.2 shows the percentage of weighted matrix values for eight variables under considerations. The chart indicates that the end-flared of the guardrail gives the highest value (87.018601). Therefore, this variable can be eliminated since it has no significant impact on the classification.

Then, data was modified. Seven variables were used in the model as input, and severity was output. The weighted matrix then was calculated. This operation was repeated to all the variables to decide which variable we have to eliminate.

Figure 5.3 presents the results of the weighted matrix which was calculated from cross validation matrix that were produced by neural networks.

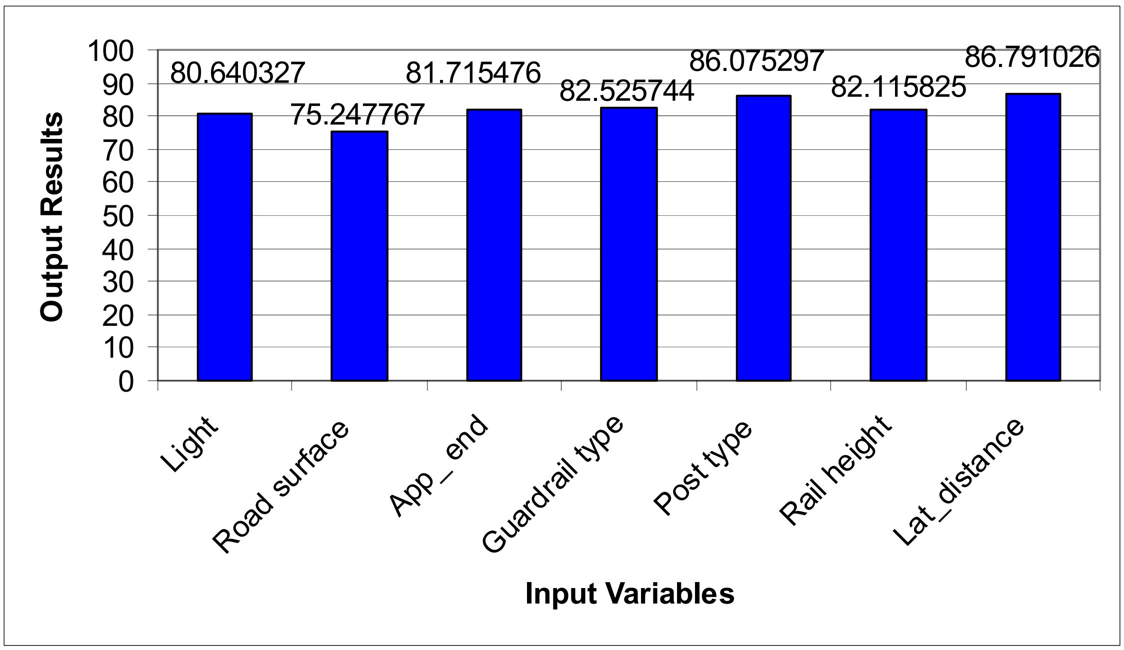

Figure 5.3 Percentage of Weighted Matrix Values for Seven Variables

From the graph, it was found that the lateral distance had no effect on severity, so this variable was deleted. The rest of variables were used in a new run. It has to be mentioned that from an engineering point of view, although lateral distance is an important factor, it seems to not be affecting the classification. Possibly, if the data included another factor it could work well to prove that the lateral distance is important.

The process was started again and the weighted matrix was calculated. The aim was to eliminate another variable. The result showed that post type has no effect on the severity. Figure 5.4 presents the weighted matrix that was calculated indicating that post type has no effect on severity.

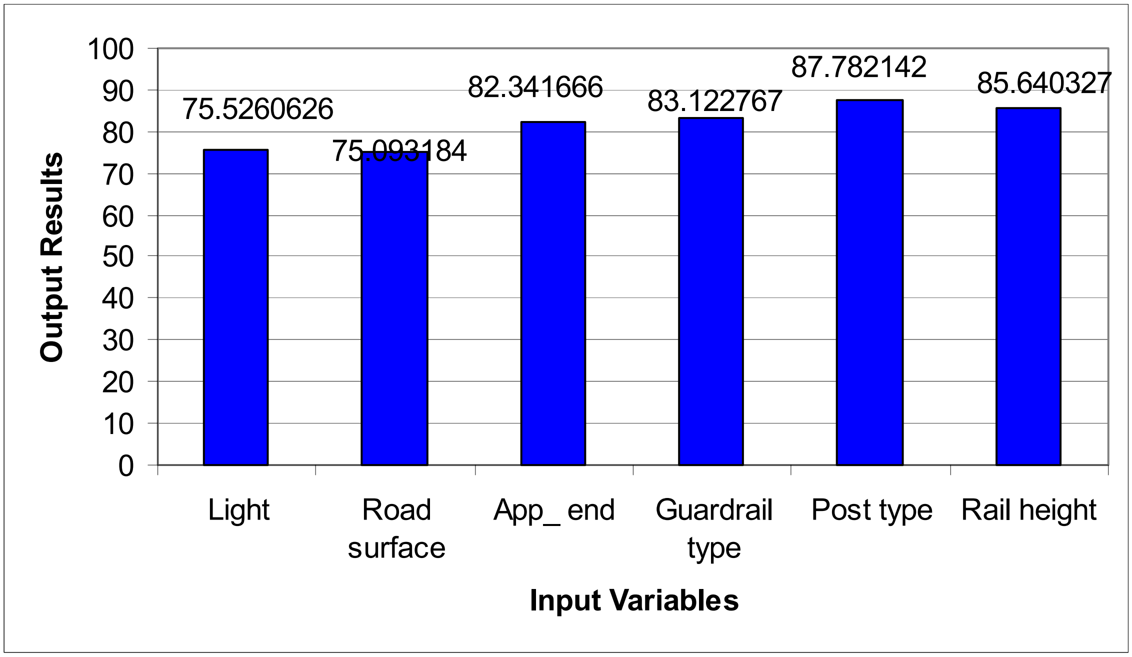

Figure 5.4 Percentage of Weighted Matrix Values for Six Variables

After eliminating post type, five variables remained. The same process was done for the remaining variables and the results are shown in Figures 5.5, 5.6 and 5.7. To reach the last results, a total of 39 Neural Networks were done. All the results can be viewed the summary presented in Table 5.4.

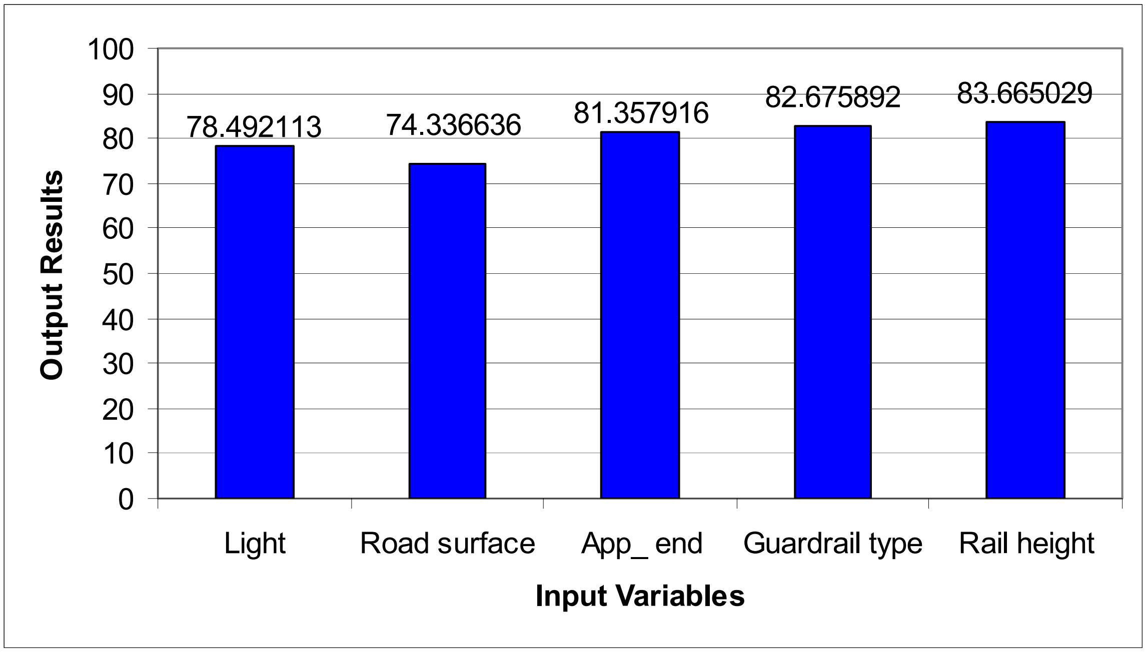

Figure 5.5 Percentage of Weighted Matrix Values for Five Variables

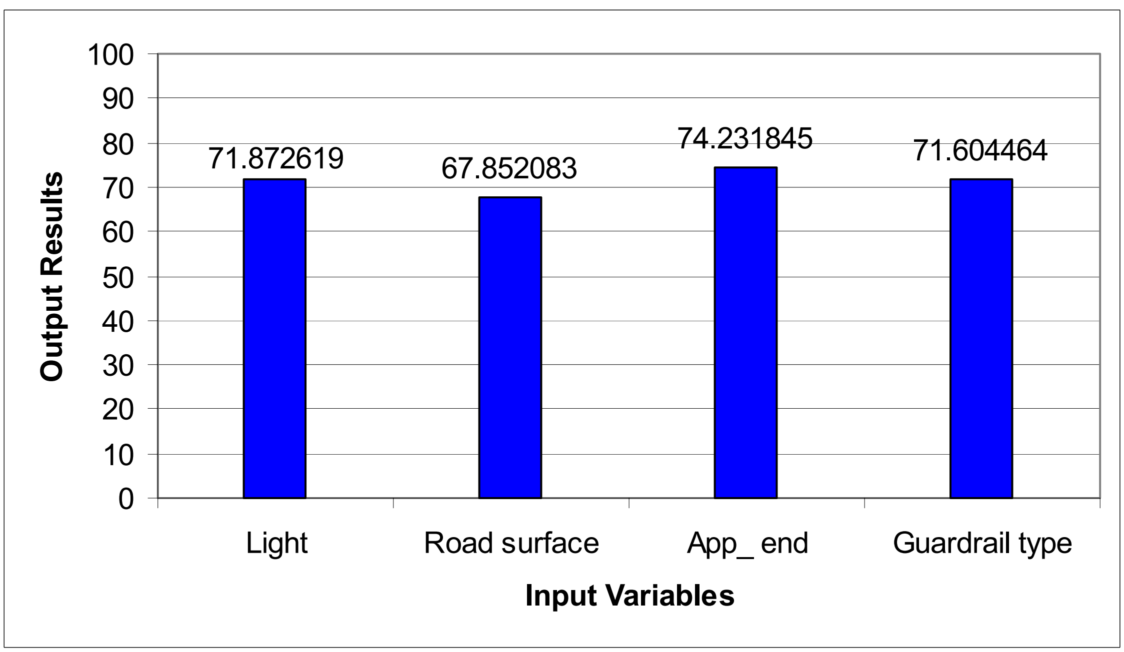

Figure 5.6 Percentage of Weighted Matrix Values for Four Variables.

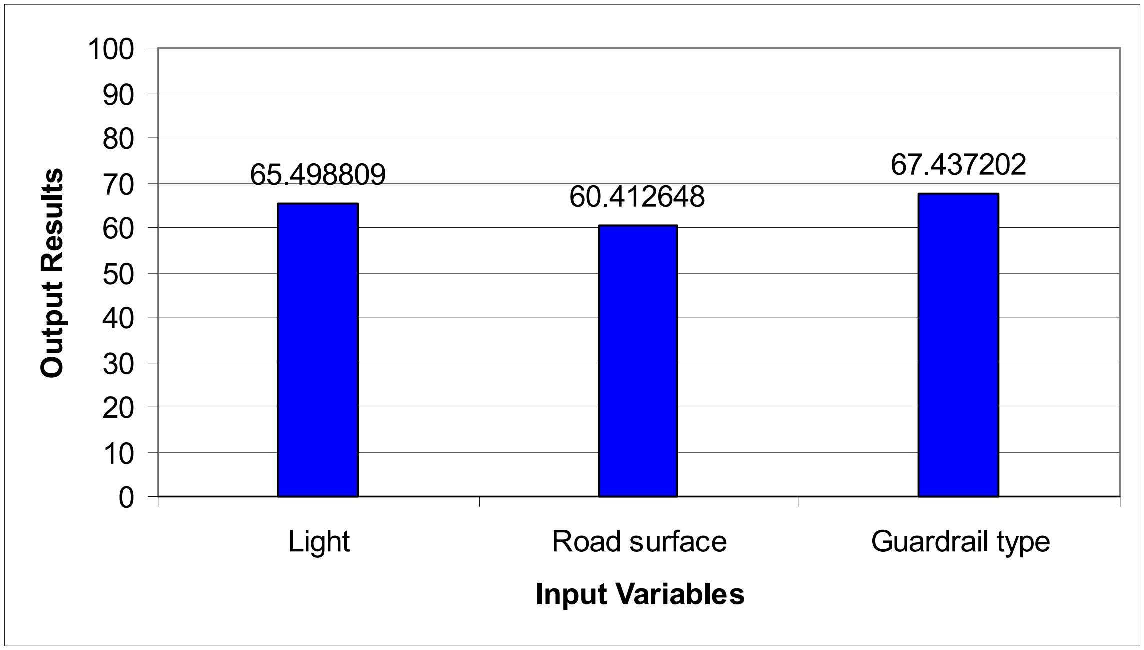

Figure 5.7 Percentage of Weighted Matrix Values for Three Variables.

Table 5.4 Summary of the Results of Road/Environment Factors

All	Variable	Light	Road surface	App-end	App-enf	Guard-type	Post -type	Rail height	Lat. Dist.
87.294672		81.240029	76.733482	85.230059	87.018601	84.51875	85.628422	82.149702	84.339434
83.3375		80.640327	75.247767	81.715476		82.525744	86.075297	82.115625	86.791026
86.791026		75.526026	75.093184	82.341666		83.122767	87.782142	85.640327	
87.782142		78.492113	74.336636	81.357916		82.675892		83.665029	
83.175		71.872619	67.852083	74.231845		71.604464			
74.358779		65.498809	60.412648			67.437202			

5.6 Input for Vehicle/Human Factors

After processes were completed with the first set of data, the second set of data, which includes the vehicle factors and human factors, was prepared for the second cluster to identify the best factors which can be used in the model. These factors are driving under influence of alcohol, speed limit, driver age, vehicle impact and number of occupants. It has to be mentioned that the driver and the vehicle have a great impact in classification of the severity. Results for training, cross validation and testing are represented Tables 5.5, 5.6 and 5.7. The result of MSE was 0.078636.

Table 5.5 Result of Training with Input Human/Vehicle in ANN

Training	2	3	4
2	82.285713	11.673470	6.040816
3	1.844168	88.197327	9.958507
4	0.602864	2.712886	96.684250

Table 5.6 Result of Cross Validation with Input Human/Vehicle in ANN

Cross validation	2	3	4
2	83.411217	10.280374	6.308411
3	1.949861	90.250694	7.799443
4	0.919540	2.988506	96.091957

Table 5.7 Result of Testing with Input Human/Vehicle in ANN

Testing	2	3	4
2	82.733810	5.995204	11.270983
3	4.807693	80.357140	14.835165
4	0.574053	1.492537	97.933411

The neural network was able to classify the different levels of severity with more than 80% accuracy. The weighted matrix was calculated and the result was 91.41. Then, the same operation was applied on vehicle/human factors to choose the best factors by eliminating variables. The results are presented in Figures 5.8, 5.9, and 5.10.

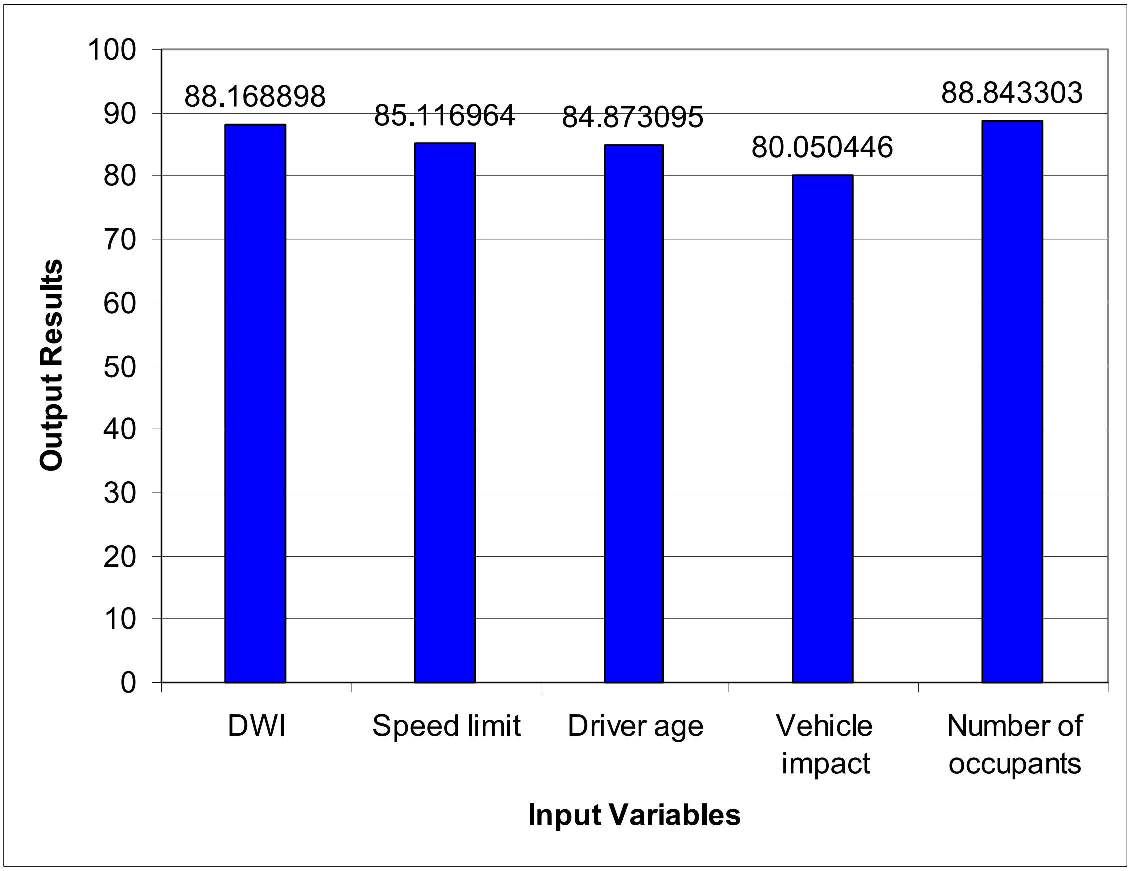

Figure 5.8 Percentage of Weighted Matrix Values for Human/Vehicle variables

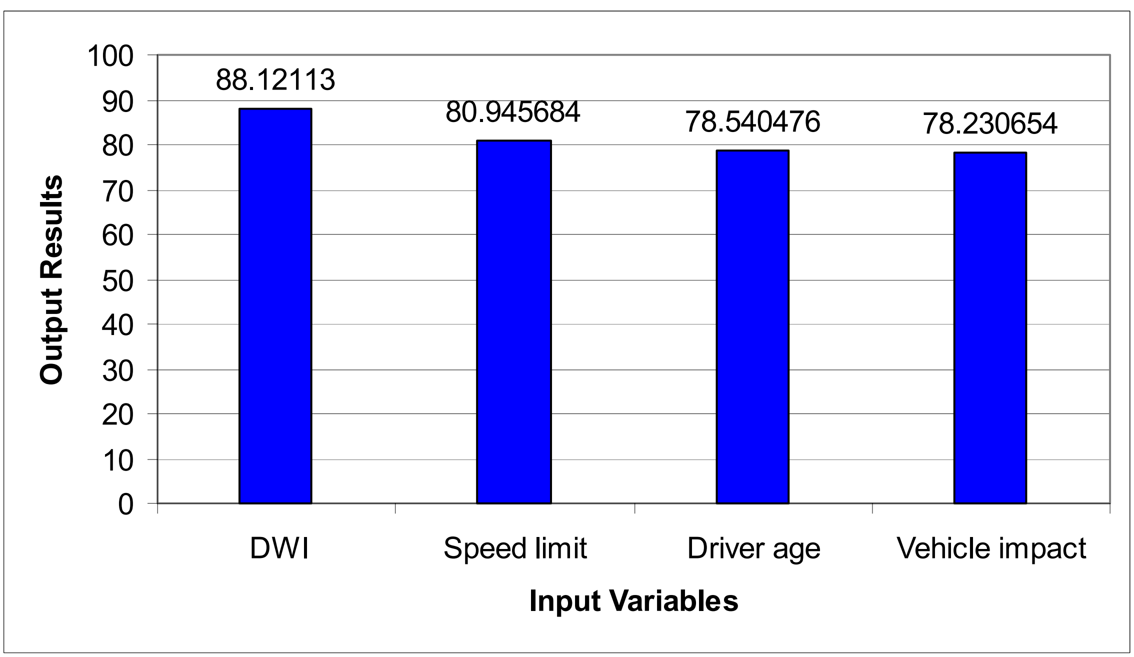

Figure 5.9 Percentage of Weighted Matrix Values for Human/vehicle variables

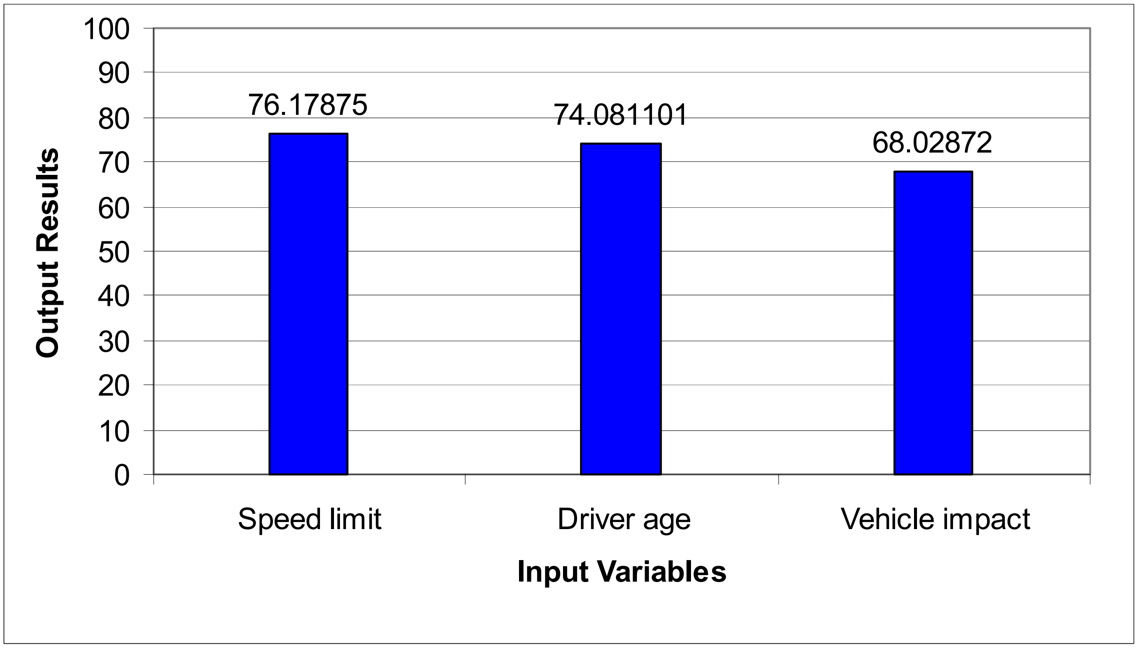

Figure 5.10 Percentage of Weighted Matrix Values for Human/vehicle of Three Variables

Three factors were eliminated in the last three figures. These factors were number of occupants, driving under influence of alcohol, and speed limit. It was found that the driver's age, speed limit, and vehicle impact have a significant effect on severity. Vehicle point of impact was found to be significant because the point of impact determines the amount of energy observed and physics of crash response. Results are summarized in Table 5.8.

Table 5.8 Summary of the Results of Human/Vehicle Factors

	DWI	Speed limit	Driver age	Vehicle impact	Numb.of occ.
91.411904	88.168898	85.116964	84.873095	80.050446	88.843303
88.508377	84.12113	80.945684	78.540476	78.230654	
84.12113		76.17875	74.081101	68.02872	

5.8 Choosing the Best Classification

Two neural networks were developed looking for best results of classification. Input variables for the first network were: Driver age, vehicle impact, light condition, road surface, and guardrail type. Input variables for the second network were: driver age, vehicle impact, light condition, road surface, end type of guardrail and the type of guardrail.

The performance of the second network was good in classification with MSE = .027089. The results of training, cross validation and testing are illustrated in Tables 5.9, 5.10, and 5.11Classification of severity Level # 2 was (93.525177%), severity Level #3 (96.428574%) and severity Level # 4 was (97.244545%), indicating very good performance of the neural network.

Table 5.9 Result of Training with Input Human/Vehicle/Roadway Variables

Training	2	3	4
2	96.261681	2.102804	1.635514
3	0.835655	96.796654	2.367688
4	0.229885	1.954023	97.816093

Table 5.10 Result of Cross Validation with Input Human/Vehicle/Roadway variables.

Cross validation	2	3	4
2	96.261681	2.102804	1.635514
3	0.835655	96.796654	2.367688
4	0.229885	1.954023	97.816093

Table 5.11 Result of Testing with Input Human/Vehicle/Roadway Variables.

Testing	2	3	4
2	93.525177	3.35314	3.117506
3	0.961538	96.428574	2.609890
4	0.459242	2.296211	97.244545

Chapter 6

Conclusions and Recommendations

6.1 Conclusions

The objective of this research has been to explore the use of Artificial Neural Networks as an analytical technique for classifying severity levels in crashes involving guardrails and to identify the most significant factors that contribute to the severity of such crashes. The major finding of this research is that ANNs performs well on both objectives. From this study the following conclusions can be made:

1. In testing the performance of the ANNs developed in this investigation, it was found that the successful classification was as follows:
   a. 93.5% for classification severity level 2 (Incapacitating injury),
   b. 96.4% for classification severity level 3 (Non-incapacitating injury), and
   c. 97.2% for classification level 4 (Possible injury).

   The above classification performance indicates that the ANNs was successful in capturing the relationship between the different variables and the severity of crash cases used in testing the performance of the network.
2. For an input of human factors (driver's age and driving under the influence of alcohol) and the vehicle factors (point of impact and speed), the ANN classification performance was 80%. This indicates that those factors have a significant effect on the classification of severity level.
3. During the process of data reduction and classification, some factors were deleted because they seemed to have little or no effect on the crash severity level. One

such factor is the lateral distance from the road edge to the guardrail. From an engineering point of view, this factor is important since the lateral distance allows room for the driver of an out of control vehicle to place it under control before the vehicle hits the guard rail.

4. Including guardrail type and the type of the guardrail end in the input of the ANNs model showed that these factors have a significant effect on the classification of crash severity.
5. Lighting condition (daylight, dawn, or darkness) is one of the important factors that influence the severity of guardrails crashes. Under poor lightening conditions, the driver's ability to respond quickly to road hazards is reduced. The driver is also less able to maneuver his vehicle to avoid hitting a guardrail especially if he cannot see it because of poor lightning condition. The ANNs developed in this study indicated the significance of road lightning condition on the severity classification performance.

6.2 Recommendations

From the study performed in this thesis, we recommend the following:

1. The present study made use of only one year of crash data because this was the only available data that could be used. However, ANNs are likely to be even more effective when larger data sets are used for training. It is recommended that all states collect crash data with guardrail inventory file and accumulate it in a national database to use for safety analysis.

2. As mentioned before, the data were collected from police crash reports that are subject to human error or missing parameters, these parameters are important in the safety analysis, so there is a need for a software package to be used in collecting the crash data properly. The package should be simple to use, with easy data entry, and logical checking to ensure that the data is as accurate as possible when entered without any missing parameters
3. The police officer decides at the seen, as much a possible, the degree of injury and puts it in the crash report form. Results of injury versus –non injury crashes will be more robust if police officers receive comprehensive training in collecting and documenting crash data on site.

List of References

1. National Council Highway Research Program (1980) “Guardrail and Median Barrier Crashworthiness” NCHRP Report No. 230

2. Ross, H. E., D. L. Sicking, and R. A. Zimmer (1993). Recommended Procedures for the Safety Performance Evaluation of Highway Features. *NCHRP Report No. 350*, National Cooperation Highway Research Program, Transportation Research Board, Washington, D.C.

3. American Association of State Highway and Transportation Officials, *Road Design Guide*, Washington, D.C.,1988.

4. Ogden, K.W. (1995). Safer Roads: A Guide to Road Safety Engineering. Ashgate Pub Co Avebury, Burlington, Vermont.

5. Lee, Jinusun and Fred L. Mannering, Analysis of Roadside Accident Frequency and Severity and Roadside Safety. *Washington State Department of Transportation Center* Report no. WA-RD 475.1 December 1999

6. Michie, Jarvis D. and Maurice E.Bronstad (1994) Highway Guardrails: Safety Features or Roadside Hazard? *Transportation Research Record* 1468,PP. 1-9

7. Jovanis, Paul P. and Hsin_Li Chang (1986) Modeling the Relationship of Accidents to Miles Traveled, *Transportation Research Record* 1068, 42-51

8. Hunter.William.W.,J.Richard.Stewart.,Kimberly.A.Eccles.,Herman.F.Huang.,Forrest.M.Council.,and David.L.Harkey,(2001) "Three-Strand Cable Median Barrier In North Carolina: An In-Service Evaluation", *Transportation Research Record* 1743, PP.97-103

9. Agent, K.R.,(1976) Guardrail performance: An Analysis of Accident Records, Commonwealth of Kentuky, DOT, Research Report No. 442, Lexington, KY, March 1976

10. Andrea Costanzo, Luigi Cicinati and Gennaro Orsi An Injury Severity Comparison On Side Impact Crashes, paper number 98_S8_W_15 http://www.nhtsa.dot.gov/esv/16198s8w15.pdf accessed on 2/16/03

11. Gattis, J.L. John P. Varghese and Larry E. Toothaker (1993) Analysis of Guardrail-end Accidents in Oklahoma, *Transportation Research Record* 1419, PP.52-62

12. Viner, John G.(1995)" Risk of Rollover in Ran-off-Road crashes", *Transportation Research Record* 1500,PP. 112-118

13. Ray, Malcolm. H. and Jennifer A. Weir (2001) “Unreported Collisions With Post-and-Beam Guardrails in Connecticut, Iowa And North Carolina”, *Transportation Research Record* 1743,PP. 111-119

14. Renski, Henry, Khattak. J. Asad and Council. M Forrest (1999) “Impact of Speed Limit Increases on Crash Injury Severity: Analysis of Single-Vehicle Crashes on North Carolina Interstate Highways” Paper represented in 78th Annual meeting Transportation Research Board Washington D.C. January 1999.

15. Council, Forrest M. and J. Richard Stewart (1996) “Severity Indexes for Roadside Objects’ *Transportation Research Record* 1528, PP.87-96.

16. Lassarre, S. (1986) The Introduction of the Variables “Traffic volume, Speed and Belt Wearing” Into A Predictive Model of the Severity of Accidents, *Accident Analysis and Prevention*, Vol. 18 No. 2,PP. 129-134

17. Jones, Ian S. and R.A. Whitfield (1988) Predicting Injury Risk with New Car Assessment Program “Crashworthiness Ratings, *Accident Analysis and Prevention*, Vol. 20, No. 6 411-419

18. Lui, K. J, Daniel McGee, Phil Rhodes. and Daniel Pollock (1988) “An Application of a Conditional Logistic Safety Belts, Principal Impact Points, and

Car Weights on Driver's Fatalities" *Journal of Safety Research*, Vol. 19, 1988, PP. 187-203

19. Kockelman,K.M., and Young-Jun Kweon (2002)"Driver Injury Severity: An Application of Ordered Probit Models" *Accident Analysis and Prevention*, Vol. 34, No: 3, pp. 313-321

20. Kim, Karl, Lawrence Nitz, James Richardson and lei Li (1995) "Analyzing the Relationship Between Crash Types and Injuries in Motor Vehicle Collisions in Hawaii", *Transportation Research Record* 1467,PP. 9-13

21. McGinnis Richard. G, Matthew J. Daves and Eric A. Hathaway (2001) Longitudinal Analysis of Fatal Run-off Road Crashes ,*Transportation Research Record* 1746,PP. 47-58

22. Shankar, Venkataraman, Fred L. Mannering and Woodrow Barfield (1996) "Statistical Analysis of Accident Severity on Rural Freeways" *Accident Analysis and Prevention*, Vol. 28, No. 3, PP.391-401

23. Mercer, G. William (1987) Influence on Passenger Vehicle Casualty Accident Frequency and Severity: Unemployment, Driver Gender, Driver Age, Drinking While Driving and Restraint Device Use, *Accident Analysis and Prevention*, vol. 19, No. 3, 231-236

24. Shibata, Akira and Katsuhiro Fukuda (1994) Risk Factors of Fatality in Motor Vehicle Traffic Accidents, *Accident Analysis and Prevention*, vol. 26, No. 3, PP. 391-397

25. Viner, John G. (1993) Harmful Events in Crashes, *Accident Analysis and Prevention*, Vol. 25 No.2 PP.139-145

26. McFadden John, Tai Yang. W, and Durrans Rocky.S (2000) "Application of Artificial Neural Networks to Predict Speeds on Two-Lane Rural Highways", *Transportation Research Record,* 1751,pp9-17

27. Abedelwahab .H and Abdel-Aty. M. A, (2001) Development of Artificial Neural Networks Models To Predict Driver Injury Severity in Traffic Accidents at Signalized Intersections, Transportation Research Board 80th annual meeting, Paper No. 01-2234 Jan. 2001 Washington D.C.

28. Sadek. A. W and Mark .C,(2003) "Modular Artificial Neural Networks for Solving the Inverse Transportation Planning Problem", Transportation Research Board 82nd Annual Meeting Washington D.C. 2003

29. Yang, H. and Fengxiang. Q(1998)"Neural Network Approach to Classification of Traffic Flow States" *Journal of Transportation Engineering*, ASCE Vol. 124,No.6,PP.521-525

30. Pant. P. D., and Purushothaman. B (1994) " Neural Network for Gap Acceptance at Stop-Controlled Intersections" *Journal of Transportation Engineering*, ASCE Vol. 120,No.3,PP.432-446

31. Ishak. S. and Haitham Al-Deek (1999) " Performance of Automatic ANN- Based Incident Detection on Freeways" *Journal of Transportation Engineering*, ASCE Vol. 125,No.4,PP.281-290

32. Teng. H., David. R. Martinelli, and Benjamin. T. (1999) " Incorporating Neural Network Traffic Prediction into Freeway Incident Detection" *Transportation Research Board*, No 1679,PP. 101-110

33. Neuro Dimensions, Inc. (2001). Neuro Solutions On-Line Manual, Version 4.1. Gainesville, Florida.

34. Principe, J.C., N. W. Euliano, and W. C. Lefebrvre. (2000). Neural and Adaptive Systems: Fundamentals through Simulations. John Wiley & Sons, Inc.

Appendix A

Michigan Crash Report Form

Authority: 1949 PA 300, Sec. 257.622
Compliance: Required
Penalty: $100 and/or 90 days

Do Not Use

Page ____ Of ____ UD-10 (1/99)

Incident #

File Class

Incident Disposition: Open / Closed — Reviewer

ORI: MI-

STATE OF MICHIGAN
Traffic Crash Report

Department Name

Crash Date (Month, Day, Year) — Crash Time (Military)

No. of Units: 1 2 3 4 5 6 7 8 9

Crash Type: Single Motor Vehicle; Head On; Head On-Left Turn; Angle; Rear End; Rear End-Left Turn; Rear End-Right Turn; Sideswipe-Same; Sideswipe-Opposite; Other/Unknown

Special Circumstances: None; Deer; School Bus; Hit and Run; Fleeing Police

Special Study: Local; State

Weather (Mark Only One): Clear; Cloudy; Fog/Smoke; Rain; Snow/Blowing Snow; Severe Wind; Sleet/Hail; Other/Unknown

Light (Mark Only One): Daylight; Dawn; Dusk; Dark-Lighted; Dark-Unlighted; Other/Unknown

Special Checks: Fatal (Report All); Corrected Copy; Replace (Entire Report); Delete (Entire Report); Non-Traffic Area; ORV/Snowmobile

County — City/Twp

Traffic Control: Signal; Stop Sign; Yield Sign; None of These

Construction Zone (if applicable) (Mark One From Each Group)
Type: Const./Maint.; Utility — Lane Closed: Yes; No — Activity: On Road; Off Road; None

LOCATION
Road Name — Divided Roadway N S E W
Distance: FT / MI — North / South — East / West — Beginning of Ramp / End of Ramp
Intersecting Road — Divided Roadway N S E W
Trafficway 1 2 3 4 — Access Control 1 2 3

Relation to Roadway (Location of First Impact): On Road; Median; Shoulder; Outside of Shoulder/Curb; Gore; Other/Unknown

Area — **Road Condition (Mark Only One):** Dry; Wet; Icy; Snowy; Muddy; Slushy; Debris; Other/Unknown

Total Lanes: 0–9 — **Speed Limit** — Posted: Yes / No

Unit Number: 1 2 3 4 5 6 7 8 9
State — Driver License Number
First Name — Middle — Last
Street Address — Phone Number
City — State — Zip
Date of Birth (Month, Day, Year)
License Type: O, CY, C, F, M, R — Sex: M, F — Hazard Action
Position — Restraint — Ambulance/Hospital
Injury: K, A, B, C, O — Ejected: Yes — Trapped: Yes — Citation Issued: Hazardous; Other
Airbag Deployed: Yes; No; Not Equipped
Unit Type: MV, P, B, E (train)
Driver Condition 1 2 3 4 5 6 7 8 9 99
Interlock: Yes / No
Refused — Not offered (Submit Results To FARS When Available)
Alcohol: Yes / No — Test Type: Field, PBT, Breath, Blood, Urine — Test Results
Drugs: Yes / No — Test Type: Blood, Urine — Test Results
Total Occup
Vehicle Registration — State — VIN — Vehicle Description (year, make, color)
Insurance — Towed To/By
Vehicle Direction: North, East, South, West — Special Vehicles 1 2 3 4 5 6
Location of Greatest Damage 0–12
Vehicle Type: PA, CY, OR, VA, MO, Other, PU, GC, Truck/Bus, ST, SM (Complete Truck/Bus Section)
Vehicle Use 1–11
First Impact — Extent of Vehicle Damage 0–7 — Driveable: Yes / No
Vehicle Defect 1–6 — Private Trailer Type 1–7

PASSENGERS
First Name — Middle — Last — Date of Birth (Month, Day, Year) — Sex: M / F — Ejected: Yes — Trapped: Yes
Street Address — Phone Number — Injury: K, A, B, C, O
City — State — Zip — Pos. — Rest. — Ambulance/Hospital — Airbag Deployed: Yes / No / Not Equipped

First Name — Middle — Last — Date of Birth (Month, Day, Year) — Sex: M / F — Ejected: Yes — Trapped: Yes
Street Address — Phone Number — Injury: K, A, B, C, O
City — State — Zip — Pos. — Rest. — Ambulance/Hospital — Airbag Deployed: Yes / No / Not Equipped

Owner / Witness / Uninjured Passenger — Name — Address — Phone Number — Age — Pos. — Rest.
Owner / Witness / Uninjured Passenger — Name — Address — Phone Number — Age — Pos. — Rest.

Person Advised of Damaged Traffic Control — Date — Time — Name — Damaged Property — Public Y / N — Owner & Phone

Do Not Write or Mark In This Area — UD-10 SERIAL NUMBER 5098571 — Serial Override Number

Do Not Write or Mark Below This Line — Do Not Write or Mark Below This Line

Unit Number ①②③④⑤⑥⑦⑧⑨ | State | Driver License Number | Date of Birth (Month, Day, Year) | License Type (O, CY, C, F, M, R) | Sex (M, F) | Hazard Action

First Name | Middle | Last | Position | Restraint | Ambulance/Hospital

Street Address | Phone Number

Unit Type: MV, P, B, E (train) | City | State | Zip | Injury: K, A, B, C, O | Ejected: Yes | Trapped: Yes | Citation Issued: Hazardous, Other

Driver Condition ①②③④⑤⑥⑦⑧⑨99 | Airbag Deployed: Yes, No, Not Equipped

Interlock: Yes, No | Refused | Not offered (Submit Results To FARS When Available)

Alcohol: Yes, No | Test Type: Field, PBT, Breath, Blood, Urine | Test Results

Drugs: Yes, No | Test Type: Blood, Urine | Test Results

Total Occup

Vehicle Registration | State | VIN | Vehicle Description (year, make, color)

Insurance | Towed To/By | Vehicle Direction: North, East, South, West | Special Vehicles ①②③④⑤⑥

Location of Greatest Damage ⓪①②③④⑤⑥⑦⑧⑨⑩⑪⑫ | Vehicle Type: PA, CY, OR, VA, MO, Other, PU, GC, Truck/Bus, ST, SM (Complete Truck/Bus Section) | Vehicle Use ①②③④⑤⑥⑦⑧⑨⑩⑪

First Impact | Extent of Vehicle Damage ⓪①②③④⑤⑥⑦ | Driveable: Yes, No | Vehicle Defect ①②③④⑤⑥ | Private Trailer Type ①②③④⑤⑥⑦

PASSENGERS

First Name | Middle | Last | Date of Birth (Month, Day, Year) | Sex: M, F | Ejected: Yes | Trapped: Yes

Street Address | Phone Number | Injury: K, B, O, A, C

City | State | Zip | Pos. | Rest. | Ambulance/Hospital | Airbag Deployed: Yes, No, Not Equipped

First Name | Middle | Last | Date of Birth (Month, Day, Year) | Sex: M, F | Ejected: Yes | Trapped: Yes

Street Address | Phone Number | Injury: K, B, O, A, C

City | State | Zip | Pos. | Rest. | Ambulance/Hospital | Airbag Deployed: Yes, No, Not Equipped

Owner, Witness, Uninjured Passenger | Name | Address | Phone Number | Age | Pos. | Rest.

Owner, Witness, Uninjured Passenger | Name | Address | Phone Number | Age | Pos. | Rest.

Unit Reported on Front

Action Prior | Sequence of Events: First, Second, Third, Fourth (⓪–⑨)

Most Harmful: Ⓜ Ⓜ Ⓜ Ⓜ

Unit Reported Above

Action Prior | Sequence of Events: First, Second, Third, Fourth (⓪–⑨)

Most Harmful: Ⓜ Ⓜ Ⓜ Ⓜ

Crash Diagram and Remarks

North

TRUCK/BUS INFORMATION

Unit No. ①②③④⑤⑥⑦⑧⑨

Carrier Name

Address

City | State | Zip

Carrier Source: Papers, Log Book, Vehicle, Driver | GVWR | Vehicle Type: AA, BB, AH, BH, AN, BN, AP, BP, AT, BX, AX, CH, AY, CP, AZ, CX, AL, Other

ICCMC | Driver's CDL Type: A, H, B, N, C, P, T, None, X | CDL Restrictions: 28, 35, 29, 36, 30

USDOT | CDL Exempt: Farm, Other

MPSC | Medical Card: Y, N

Interstate, Intra (MI Only)

Type & Axles Per Unit: First, Second, Third, Fourth | Hazardous Material: Placard Y N, Cargo Spill Y N, ID #, Class #

Cargo Body Type ①②③④⑤⑥⑦⑧

UD-10 SERIAL NUMBER 5098572 | Investigated at Scene Ⓨ Ⓝ | Reported Date/Time | Investigator Name(s) & Badge # (Print Only)

Forward Original To: Michigan State Police, Traffic Crash Reporting Section, 7150 Harris Drive, Lansing, MI 48913

Do Not Write or Mark On This Side of The Line

Do Not Write or Mark Below This Line

Appendix B

Results of training and cross validation metrics

First: Road/Environment
(1)

Input	Road surface	App-End	App-Enf	Guardrail type	Post type	Rail height	Lateral distance

Training	2	3	4
2	67.020409	19.918367	13.061225
3	0.783771	92.485016	6.731213
4	2.411454	13.564431	84.024117

Cross Validation	2	3	4
2	65.420563	17.056074	17.523365
3	1.114206	88.857941	10.027855
4	3.333333	14.252873	82.413795

(2)

Input	Light condition	App-End	App-Enf	Guardrail type	Post type	Rail height	Lateral distance

Training	2	3	4
2	73.224487	6.775510	20.0000
3	5.855233	78.699860	15.444905
4	14.845516	7.347400	77.807083

Cross Validation	2	3	4
2	81.542053	1.869159	16.588785
3	8.774373	78.272980	12.952646
4	20.0000	6.781609	73.218391

(3)

Input	Light condition	Road surface	App-Enf	Guardrail type	Post type	Rail height	Lateral distance

Training	2	3	4
2	70.448982	11.428572	18.122450
3	3.411711	89.764870	6.823421
4	2.486812	6.895252	90.617935

Cross Validation	2	3	4
2	72.196259	6.542056	21.261683
3	4.596100	87.465179	7.938719
4	2.413793	8.045977	89.540230

(4)

Input	Light condition	Road surface	App-End	Guardrail type	Post type	Rail height	Lateral distance

Training	2	3	4
2	69.224487	7.020408	23.755102
3	0.599355	90.640846	8.759797
4	2.712886	4.973625	92.313492

Cross Validation	2	3	4
2	69.859710	4.672897	25.467289
3	0.139276	90.389969	9.470752
4	2.643678	5.057471	92.298851

(5)

Input	Light condition	Road surface	App-End	App-Enf	Post type	Rail height	Lateral distance

Training	2	3	4
2	62.183674	9.061225	27.755102
3	2.120793	84.555092	13.324113
4	1.469480	4.483798	94.046722

Cross Validation	2	3	4
2	60.514019	7.242990	32.242992
3	0.835655	84.818939	14.345404
4	1.379310	2.988506	95.632187

(6)

Input	Light condition	Road surface	App-End	App-Enf	Guardrail type	Rail height	Lateral distance

Training	2	3	4
2	77.714287	4.734694	17.551020
3	1.152605	82.388199	16.459198
4	3.240392	4.10708	92.652603

Cross Validation	2	3	4
2	77.803741	3.504673	18.691589
3	1.114206	82.033424	16.852367
4	2.988506	4.712644	92.298851

(7)

Input	Light condition	Road surface	App-End	App-Enf	Guardrail type	Post type	Lateral distance

Training	2	3	4
2	70.775513	18.775511	10.448979
3	2.858460	86.583679	10.557860
4	2.486812	10.926903	86.586282

Cross Validation	2	3	4
2	66.121498	18.691589	15.186916
3	2.089127	85.236771	12.674095
4	1.494253	11.264368	87.241379

(8)

Input	Light condition	Road surface	App-End	App-Enf	Guardrail type	Post type	Rail height

Training	2	3	4
2	76.326530	7.020408	16.653061
3	3.596127	83.310280	13.093592
4	3.579503	4.709872	91.710625

Cross Validation	2	3	4
2	73.364487	6.308411	20.327103
3	3.064067	83.565460	13.370474
4	4.022988	5.862069	90.114944

(9) All Factors after eliminating App- enf

Input	Light condition	Road surface	App-End	Guardrail type	Post type	Rail height	Lateral distance

Training	2	3	4
2	64.816330	11.510204	23.673470
3	1.383126	87.229141	11.387736
4	2.637528	5.124341	92.238129

Cross Validation	2	3	4
2	61.682243	7.476635	30.841122
3	0.696379	86.350975	12.952646
4	2.873563	6.436781	90.689651

(10)

Input	Road surface	App-End	Guardrail type	Post type	Rail height	Lateral distance

Training	2	3	4
2	66.285713	21.306122	12.408163
3	0.691563	92.300598	7.007838
4	2.901281	16.390354	80.708366

Cross Validation	2	3	4
2	63.551403	20.327103	16.121496
3	0.696379	93.036209	6.267409
4	2.988506	18.505747	78.505745

(11)

Input	Light condition	App-End	Guardrail type	Post type	Rail height	Lateral distance

Training	2	3	4
2	60.081635	11.020409	28.897959
3	1.244813	78.192719	20.562471
4	11.793519	8.703843	79.502640

Cross Validation	2	3	4
2	66.822433	7.009346	26.168224
3	1.114206	79.944290	18.941504
4	14.137931	10.459770	75.402298

(12)

Input	Light condition	Road surface	Guardrail type	Post type	Rail height	Lateral distance

Training	2	3	4
2	49.142857	13.061225	37.795918
3	0.875980	86.721992	12.402029
4	1.657875	6.141673	92.200455

Cross Validation	2	3	4
2	50.000	12.149532	37.850468
3	0.835655	86.350975	12.813371
4	0.919540	6.321839	92.758621

(13)

Input	Light condition	Road surface	App-End	Post type	Rail height	Lateral distance

Training	2	3	4
2	66.857140	4.408163	28.734694
3	5.394191	80.497925	14.107883
4	1.620196	5.011304	93.368500

Cross Validation	2	3	4
2	64.485985	2.570093	32.943924
3	4.456824	80.919220	14.623956
4	1.379310	6.206897	92.413795

(14)

Input	Light condition	Road surface	App-End	Guardrail type	Rail height	Lateral distance

Training	2	3	4
2	72.326530	10.285714	17.387754
3	0.461042	92.669434	6.869525
4	2.637528	6.631500	90.730972

Cross Validation	2	3	4
2	69.158882	9.345795	21.495327
3	0.557103	88.857941	10.584958
4	1.724138	6.551724	91.724136

(15)

Input	Light condition	Road surface	App-End	Guardrail type	Post type	Lateral distance

Training	2	3	4
2	66.612244	13.632653	19.755102
3	2.305210	81.143387	16.551407
4	2.411454	6.556142	91.032402

Cross Validation	2	3	4
2	62.383179	10.514019	27.102804
3	1.810585	79.526459	18.662952
4	1.034483	5.287356	93.678162

(16)

Input	Light condition	Road surface	App-End	Guardrail type	Post type	Rail height

Training	2	3	4
2	78.040817	8.897959	13.061225
3	2.305210	92.070076	5.624712
4	3.579503	7.121326	89.299171

Cross Validation	2	3	4
2	75.700935	8.644860	15.654205
3	2.506964	89.972145	7.520891
4	2.873563	7.701149	89.425285

(17) All Variables after Eliminating Lateral distance

Input	Light condition	Road surface	App-End	Guardrail type	Post type	Rail height

Training	2	3	4
2	78.040817	8.897959	13.061225
3	2.305210	92.070076	5.624712
4	3.579503	7.121326	89.299171

Cross Validation	2	3	4
2	75.700935	8.644860	15.654205
3	2.506964	89.972145	7.520891
4	2.873563	7.701149	89.425285

(18)

Input	Road surface	App-End	Guardrail type	Post type	Rail height

Training	2	3	4
2	52.816326	24.489796	22.693878
3	0.645459	88.151222	11.203320
4	1.733233	17.746798	80.519966

Cross Validation	2	3	4
2	50.934578	17.757010	31.308411
3	0.696379	81.754875	17.548746
4	1.609195	16.436781	81.954025

(19)

Input	Light condition	App-End	Guardrail type	Post type	Rail height

Training	2	3	4
2	65.714287	13.387755	20.897959
3	1.383126	82.526512	16.090364
4	13.677468	13.187641	73.134888

Cross Validation	2	3	4
2	72.897194	11.682243	15.420561
3	0.974930	85.097496	13.927577
4	16.206896	15.862069	67.931038

(20)

Input	Light condition	Road surface	Guardrail type	Post type	Rail height

Training	2	3	4
2	59.836735	11.346939	28.816326
3	2.443522	84.140160	13.416321
4	4.107008	8.138659	87.754333

Cross Validation	2	3	4
2	59.112148	7.242990	33.644859
3	0.557103	83.286911	16.155989
4	2.068965	5.402299	92.528732

(21)

Input	Light condition	Road surface	App-End	Post type	Rail height

Training	2	3	4
2	74.612244	5.551021	19.836735
3	4.933149	82.710930	12.355925
4	3.956292	6.706858	89.336853

Cross Validation	2	3	4
2	69.392525	4.906542	25.700935
3	3.064067	84.261841	12.674095
4	2.758621	8.505747	88.735634

(22)

Input	Light condition	Road surface	App-End	Guardrail type	Rail height

Training	2	3	4
2	78.285713	5.959184	15.755102
3	1.936376	87.275246	10.788382
4	2.373775	5.162020	92.464203

Cross Validation	2	3	4
2	75.233643	4.672897	20.093458
3	1.810585	88.440109	9.749304
4	2.183908	4.597701	93.218391

(23)

Input	Light condition	Road surface	App-End	Guardrail type	Post type

Training	2	3	4
2	77.142860	7.428571	15.428572
3	3.550023	84.094055	12.355925
4	3.504145	7.385079	89.110779

Cross Validation	2	3	4
2	77.803741	5.607477	16.588785
3	3.064067	84.122566	12.813371
4	3.448276	5.977012	90.574715

(24) All Variables after Eliminating Post Type

Input	Light condition	Road surface	App-End	Guardrail type	Rail height

Training	2	3	4
2	78.285713	5.959184	15.755102
3	1.936376	87.275246	10.788382
4	2.373775	5.162020	92.464203

Cross Validation	2	3	4
2	75.233643	4.672897	20.093458
3	1.810585	88.440109	9.749304
4	2.183908	4.597701	93.218391

(25)

Input	Road surface	App-End	Guardrail type	Rail height

Training	2	3	4
2	59.346939	26.693878	13.959184
3	0.783771	89.488243	9.727985
4	2.260739	17.445366	80.293900

Cross Validation	2	3	4
2	58.177570	23.598131	18.224298
3	0.974930	88.440109	10.584958
4	2.183908	17.931034	79.885056

(26)

Input	Light condition	App-End	Guardrail type	Rail height

Training	2	3	4
2	64.653061	11.591837	23.755102
3	2.305210	76.579071	21.115721
4	14.770158	12.773172	72.456673

Cross Validation	2	3	4
2	73.130844	8.177570	18.691589
3	0.974930	82.729805	16.295265
4	16.551723	15.402299	68.045975

(27)

Input	Light condition	Road surface	Guardrail type	Rail height

Training	2	3	4
2	64.816330	12.081633	23.102041
3	1.890272	84.555092	13.554633
4	4.822909	8.440090	86.737000

Cross Validation	2	3	4
2	65.654205	7.943925	26.401869
3	2.506964	82.172699	15.320334
4	5.057471	6.896552	88.045975

(28)

Input	Light condition	Road surface	App-End	Rail height

Training	2	3	4
2	68.571426	4.906542	26.448980
3	2.996773	79.299217	17.704012
4	1.582517	5.614167	92.803314

Cross Validation	2	3	4
2	66.121498	4.906542	28.971962
3	2.785515	78.412254	18.802229
4	1.494253	4.482759	94.022987

(29)

Input	Light condition	Road surface	App-End	Guardrail type

Training	2	3	4
2	72.653061	8.816326	18.530613
3	3.227294	81.420006	15.352697
4	4.446119	7.008289	88.545593

Cross Validation	2	3	4
2	68.925232	6.775701	24.299065
3	2.506964	81.476326	16.016712
4	1.494253	6.091954	92.413795

(30) All Variables after Eliminating Rail Height

Input	Light condition	Road surface	App-End	Guardrail type

Training	2	3	4
2	69.061226	8.163265	22.775511
3	2.166897	82.388199	15.444905
4	2.373775	7.535795	90.090431

Cross Validation	2	3	4
2	67.523361	7.943925	24.532711
3	2.228412	81.337044	16.434540
4	1.839080	6.091954	92.068962

(31)

Input	Road surface	App-End	Guardrail type

Training	2	3	4
2	59.102039	24.816326	16.081633
3	3.135085	81.327805	15.537114
4	4.521477	22.155237	73.323282

Cross Validation	2	3	4
2	57.710281	23.831776	18.457945
3	2.228412	81.894150	15.877438
4	4.942529	24.712645	70.344826

(32)

Input	Light condition	App-End	Guardrail type

Training	2	3	4
2	56.734695	10.530612	32.734695
3	4.518211	75.472565	20.009220
4	15.523738	16.164280	68.311981

Cross Validation	2	3	4
2	73.364487	5.841122	20.794392
3	9.888579	74.094711	16.016712
4	22.643679	17.241379	60.114941

(33)

Input	Light condition	Road surface	Guardrail type

Training	2	3	4
2	56.816326	13.877551	29.306122
3	4.057169	67.680962	28.261871
4	4.182366	8.854559	86.963074

Cross Validation	2	3	4
2	53.971962	13.084112	32.943924
3	3.899721	69.498604	26.601671
4	2.873563	9.425287	87.701149

(34)

Input	Light condition	Road surface	App-End

Training	2	3	4
2	35.265305	17.959183	46.775509
3	4.379899	72.936836	22.683264
4	3.315750	8.929917	87.754333

Cross Validation	2	3	4
2	33.644859	11.915888	54.439251
3	4.456824	75.487465	20.055710
4	3.103448	10.574713	86.321838

(35) All Variables after Eliminating App End

Input	Light condition	Road surface	Guardrail type

Training	2	3	4
2	54.612244	17.306122	28.081633
3	3.411711	69.571228	27.017059
4	4.408440	10.550113	85.041451

Cross Validation	2	3	4
2	51.168224	15.887851	32.943924
3	1.949861	71.448471	26.601671
4	2.643678	9.655172	87.701149

(36)

Input	Road surface	Guardrail type

Training	2	3	4
2	47.510204	34.285713	18.204082
3	2.166897	77.823883	20.009220
4	4.107008	30.972118	64.920876

Cross Validation	2	3	4
2	47.897198	33.411217	18.691589
3	1.253482	72.980499	25.766016
4	2.873563	29.425287	67.701149

(37)

Input	Light condition	Guardrail type

Training	2	3	4
2	44.73495	11.346939	43.918365
3	4.840940	60.811436	34.347626
4	12.999247	17.370008	69.630745

Cross Validation	2	3	4
2	44.626167	10.981308	44.392525
3	4.038997	66.155991	29.805014
4	13.563218	23.333334	63.103447

(38)

Input	Light condition	Road surface

Training	2	3	4
2	50.12248	8.408163	41.469387
3	8.944214	60.488705	30.567081
4	6.518463	9.984928	83.496613

Cross Validation	2	3	4
2	47.429905	9.579439	42.990654
3	10.584958	62.256268	27.158773
4	7.586207	11.264368	81.149422

Second Human/Vehicle
(1)

Input	D w i	Speed limit	Driver age	Vehicle impact	Number of Occupants

Training	2	3	4
2	82.285713	11.673470	6.040816
3	1.844168	88.197327	9.958507
4	0.602864	2.712886	96.684250

Cross Validation	2	3	4
2	83.411217	10.280374	6.308411
3	1.949861	90.250694	7.799443
4	0.919540	2.988506	96.091957

(2)

Input	Speed limit	Driver age	Vehicle impact	Number of Occupants

Training	2	3	4
2	73.061226	14.448979	12.489796
3	0.0	85.615494	14.384509
4	0.263753	4.634514	95.101730

Cross Validation	2	3	4
2	71.728973	13.551402	14.719626
3	0.0	88.161560	11.838440
4	0.689655	3.333333	95.977013

(3)

Input	D w i	Driver age	Vehicle impact	Number of Occupants

Training	2	3	4
2	81.306122	11.020409	7.673470
3	5.117566	77.132317	17.750114
4	0.979653	7.686511	91.333839

Cross Validation	2	3	4
2	81.074768	9.813084	9.112149
3	4.874652	79.944290	15.181059
4	0.114943	8.620689	91.264366

(4)

Input	D w i	Speed limit	Vehicle impact	Number of Occupants

Training	2	3	4
2	77.142860	12.897959	9.959184
3	4.103273	69.064087	26.832642
4	0.414469	2.185380	97.400154

Cross Validation	2	3	4
2	78.738319	13.317757	7.943925
3	4.038997	72.423401	23.537605
4	0.459770	1.494253	98.045975

(5)

Input	D w i	Speed limit	Driver age	Number of Occupants

Training	2	3	4
2	61.877552	20.163265	17.959183
3	0.0	72.153069	27.846933
4	0.0	7.611153	92.388847

Cross Validation	2	3	4
2	62.383179	19.626167	17.990654
3	0.0	74.930359	25.069637
4	0.0	7.356322	92.64367

(6)

Input	D w i	Speed limit	Driver age	Vehicle impact

Training	2	3	4
2	82.367348	8.244898	9.387755
3	4.702628	81.143387	14.153988
4	0.226074	2.675207	97.098717

Cross Validation	2	3	4
2	90.420563	1.635514	7.943925
3	8.635098	78.412254	12.952646
4	0.459770	2.873563	96.666664

(7) All Variables after Eliminating Number of Occupants

Input	D W i	Speed limit	Driver age	Vehicle impact

Training	2	3	4
2	79.591835	11.428572	8.979591
3	2.904564	79.990776	17.104656
4	0.715901	3.202713	96.081390

Cross Validation	2	3	4
2	81.308411	11.682243	7.009346
3	2.089137	82.311981	15.598886
4	0.459770	2.528736	97.011497

(8)

Input	Speed limit	Driver age	Vehicle impact

Training	2	3	4
2	82.285713	9.795918	7.918367
3	4.794837	78.653755	16.551407
4	1.281085	8.703843	90.015068

Cross Validation	2	3	4
2	85.046730	7.943925	7.009346
3	6.267409	77.158775	16.573816
4	0.689655	9.885057	89.425285

(9)

Input	D W i	Driver age	Vehicle impact

Training	2	3	4
2	75.510201	7.591837	16.897959
3	7.330567	72.752419	19.917013
4	6.857574	5.990957	87.151466

Cross Validation	2	3	4
2	82.242989	2.570093	15.186916
3	10.445683	70.752090	18.802229
4	5.862069	5.402299	88.735634

(10)

Input	D W i	Speed limit	Vehicle impact

Training	2	3	4
2	69.877548	16.571428	13.551021
3	4.933149	66.528351	28.538498
4	1.356443	10.700829	87.942726

Cross Validation	2	3	4
2	74.065422	14.252337	11.682243
3	5.153203	66.295265	28.551533
4	0.804598	8.505747	90.689651

(11)

Input	D W i	Speed limit	Driver age

Training	2	3	4
2	63.591835	18.938776	17.469387
3	1.290918	69.110191	29.598894
4	1.959307	9.608139	88.432556

Cross Validation	2	3	4
2	64.018692	21.728971	14.252337
3	1.392758	72.841225	25.766016
4	1.494253	9.080460	89.425285

(12) All Variables after Eliminating D W I

Input	Speed limit	Driver age	Vehicle impact

Training	2	3	4
2	82.285713	9.795918	7.918367
3	4.794837	78.653755	16.551407
4	1.281085	8.703843	90.015068

Cross Validation	2	3	4
2	85.046730	7.943925	7.009346
3	6.267409	77.158775	16.573816
4	0.689655	9.885057	89.425285

(13)

Input	Driver age	Vehicle impact

Training	2	3	4
2	87.918365	4.081633	8.0
3	13.785154	66.159523	20.055325
4	14.204973	9.834212	75.960815

Cross Validation	2	3	4
2	91.355141	2.336449	6.308411
3	15.877438	65.181061	18.941504
4	11.954023	10.0	78.045975

(14)

Input	Speed limit	Vehicle impact

Training	2	3	4
2	58.367348	24.897959	16.734694
3	0.599355	71.692024	27.708622
4	0.791258	18.538055	80.670685

Cross Validation	2	3	4
2	60.280373	22.663551	17.056074
3	0.0	76.601669	23.398329
4	0.229885	21.264368	78.505745

(15)

Input	Speed limit	Driver age

Training	2	3	4
2	52.734695	11.510204	35.755100
3	0.0	54.402950	45.597050
4	0.0	10.700829	89.299171

Cross Validation	2	3	4
2	53.738319	9.112149	37.149532
3	0.0	49.442898	50.557102
4	0.0	9.885057	90.114944

Five variables

Input	Driver age	Vehicle impact	Light condition	Road surface	Guardrail type

Training	2	3	4
2	83.346939	9.387755	7.265306
3	1.429230	92.761642	5.809129
4	0.0	3.202713	96.797287

Cross Validation	2	3	4
2	85.747665	5.841122	8.411215
3	1.392758	91.086349	7.520891
4	0.0	3.678161	96.321838

Best Results With Six Variables

Input	Driver age	Vehicle impact	Light condition	Road surface	App- End	Guardrail type

Training	2	3	4
2	96.261681	2.102804	1.635514
3	0.835655	96.796654	2.36788
4	0.229885	1.954023	97.816093

Cross Validation	2	3	4
2	96.261681	2.102804	1.635514
3	0.835655	96.796654	2.367688
4	0.229885	1.954023	97.816093