ANALYSIS AND MODELLING FOR RISK MANAGEMENT FOR
UNDERGROUND COAL MINES' SAFETY

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ÖZLEM DENİZ ERATAK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
MINING ENGINEERING

JANUARY 2014

Approval of the thesis:

ANALYSIS AND MODELLING FOR RISK MANAGEMENT FOR UNDERGROUND COAL MINES' SAFETY

submitted by **ÖZLEM DENİZ ERATAK** in partial fulfilment of the requirements for the degree of **Doctor of Philosophy in Mining Engineering, Middle East Technical University** by,

Prof. Dr. Canan Özgen _______________________
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Ali İhsan Arol _______________________
Head of Department, **Mining Engineering**

Prof Dr. Celal Karpuz _______________________
Supervisor, **Mining Engineering Department, METU**

Examining Committee Members:

Prof Dr. Tevfik Güyagüler _______________________
Mining Engineering Department, METU

Prof Dr. Celal Karpuz _______________________
Mining Engineering Department, METU

Prof Dr. Murat Gündüz _______________________
Civil Engineering Department, METU

Assoc. Prof. Dr. Yasin Dursun Sarı _______________________
Civil Engineering Department, Atılım University

Assoc.. Prof. Dr. Nuray Demirel _______________________
Mining Engineering Department, METU

Date: _______________________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: ÖZLEM DENİZ ERATAK

Signature

ABSTRACT

ANALYSIS AND MODELLING FOR RISK MANAGEMENT FOR UNDERGROUND COAL MINES' SAFETY

Eratak, Özlem Deniz
Ph.D, Department of Mining Engineering
Supervisor: Prof. Dr Celal Karpuz
January, 2014, 156 pages

Safety in underground coal mining has become an important issue because of increasing number of accidents. There are many different hazards may cause these accidents and the most efficient method for coping with risks is the use of risk management techniques.

In this thesis, accident data including workday losses, age of the injured, organ affected by accident, season, shift and reason of the accident was collected from TKI mines (ELI- Soma Eynez and GLI Tuncbilek) and TTK mines. Those variables were initially analysed by using basic statistics to have the general information about the most hazardous conditions. Comparison was made between these mines. Then, a risk analysis study was performed using severity, probability and exposure components. Risk matrices were developed and the most hazardous places were determined together with comparison of those three mines. Probability analysis was performed to understand the expected accident frequencies in each mine and reliability in a time period between accidents.

The study was also targeted to develop a model for severity component using three different methods which are regression, neural network and fuzzy logic techniques. These techniques applied to every mines data and decision analysis was made to choose the most suitable technique by comparing the results. Finally, future accident estimation models were developed with regression and neural network techniques based on the data such as number of accidents, deaths, injured, total working hours, total workers and total raw coal production of those mines.

Key Words: Risk analysis, risk management, coal mining, neural network, regression, fuzzy logic

ÖZ

YERALTI KÖMÜR MADENCİLİĞİNDE GÜVENLİK İÇİN RİSK YÖNETİMDE ANALİZ VE MODELLEME

Eratak, Özlem Deniz
Doktora, Maden Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr Celal Karpuz
Ocak, 2014, 156 sayfa

Yeraltı kömür madenciliğinde iş güvenliği; kaza sayılarının artmasıyla birlikte oldukça önemli bir konu haline gelmiştir. Bu kazalara bir çok değişik tehlike neden olmakla birlikte, riskler ile başedebilmenin en etkili yolu risk yönetimi teknikleridir.

Bu tezde, iş günü kayıpları, kazaya uğrayan kişinin yaşı ve hangi organının etkilendiği, kazanın gerçekleştiği mevsim, vardiya ve kazanın nedeni gibi kaza verileri, TKI madenleri (ELI-Soma Eynez ve GLI Tuncbilek) ile TTK madenlerinden toplanmıştır. İlk olarak, değişkenler basit istatistik yöntemleriyle analiz edilip en tehlikeli durumlar hakkında genel bilgi edinilmiştir. Daha sonra risk analizi çalışması yapılmış olup, çalışmada kazaların şiddeti, olasılığı ve maruziyet bileşenleri dikkate alınmıştır. Buna istinaden risk matrisleri oluşturulmuş ve en tehlikeli alanlar, üç madenin de karşılaştırılmasıyla belirlenmiştir. Her bir madendeki beklenen kaza sıklığı ve kazalar arasında geçen zaman belirlenmek üzere olasılık analizi gerçekleştirilmiştir.

Ayrıca çalışma, şiddet bileşenini belirlemek için regresyon, sinir ağları ve bulanık mantık teknikleriyle model oluşturmayı hedeflemiştir. Bu teknikler, bütün maden verileri üzerinde uygulanarak, karşılaştırma analiziyle en uygun teknik belirlenmiştir. Son olarak, gelecek kaza tahmin modeli regresyon ve sinir ağları teknikleriyle, her

bir madene ait kaza sayıları, ölümler, yaralı sayıları, toplam çalışma saatleri ve toplam ham kömür üretimi verileriyle gerçekleştirilmiştir.

Anahtar Kelimeler: Risk analizi, risk yönetimi kömür madenciliği, sinir ağları, regresyon, bulanık mantık

ACKNOWLEDGMENTS

First, I would like to thank my supervisor Prof.Dr.Celal Karpuz for his guidance, support and patience. It was a great honour for me to work with him in this study. I also would like to thank Prof.Dr. Murat Gündüz, Prof. Dr. Yasin Dursun Sarı and Prof.Dr. Tevfik Güyagüler for their valuable feedbacks and time.

I would like to give my special thanks to my family without their encouragement I could never achieve this.

Finally, I would like to thank my dear friend Hidayet Doğan, who always supports and helps in my desperate times until the end of this thesis. I will never forget his great contributions that give me perspective.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

Though the history of coal extraction dates back to at least 287 B.C. (Oleson, 2008), its consumption increased very rapidly in 18th century because of its requirement in steam engines during industrial revolution. As a result of this increasing demand pressure, more and more mining sites were getting developed for commercial purpose, and new machines were also invented and used for coal extraction .Underground mining is relatively more sophisticated as compared to surface mining, as it can be used to recover coal from much deeper layers of earth. This method was being developed during industrial revolution, and was found to be more hazardous than conventional surface mining method (Langton, 1979). All these resulted in poor working conditions for industry labours and the scenario further worsened after the discovery of assembly line manufacturing process. Such methods resulted in improved production at the expense of labours working in dangerous working conditions for long hours, and increased frequency of such accidents (Karmis, 2001).

Several authors have described accidents since industrial revolution. For instance, (Richards, 2007) provided details on nine coal mine accidents in Pennsylvania States. There are numerous latest examples of major coal-mine accidents as well. For example, an accident as recent as that of December 2009, Sebnem, 2009 reported the accident of an explosion in underground coal mine due to build-up of methane gas, which caused a fire claiming the lives of 19 workers in Bursa, Turkey. This was followed by implementation of tighter rules and safety regulations for industrial labours of Turkey (Gozuyilmaz, 2003). This benefitted underground coal mines, where death rates were more than that in any other sector employing industrial labours.

These industrial developments and underground coal mine accidents motivated the research work for discovery and identification of methods to remove them. There are several approaches of analysing the reasons of such accidents. Ericson (2005) explained how safety risks are related to accidents, and suggested several types of hazard analysis techniques, including both qualitative and quantitative methods. Qualitative methods are used for exploration and identification purpose, but quantitative methods provide further insights on other aspects, such as relative effect of accidents on causes. Some past research works on quantitative methods include that statistical analysis for development of a relationship between such accidents and causes, and the analysis of the extent to which they affect the probability of accident. (McDonald et.al. 1980) Major technical improvements and research work done on similar projects since 1980 have opened up opportunities for statistical analysis and modelling of underground coal mines using new and sophisticated methods.

1.1 Scope of the Study

Understanding risk is a general difficulty in mining sector and it is very important to foresee the accidents using the previous accident statistics.

Since risk is derived from severity, probability and exposure of accidents (or a combination of both), physical quantities representing these two may be considered as the "result quantities" of statistical analysis. For instance, probability may be represented by number of accidents corresponding to a particular parameter. Similarly, severity may be represented by damage in the form of number of workdays lost, injuries, deaths or disability cases per accident, and the product of severity and frequency may be represented by cumulative results, such as overall number of injuries, deaths or disability cases throughout the history. These all may be considered as output parameters while the cause of accident may be linked to a reason of accident or type of hazard. However, there are several other parameters directly or indirectly affecting the outputs, such as organ affected by accident, working shift (day/evening/night), time of the year or even age of worker. Some inputs like reason of damage and affected organ directly affect the injured person while the effects of age, time of year or work shift may be assumed to affect them indirectly.

The general aim of the study is to develop a relationship between accidents and their consequences in two lignite mines of TKI (ELI-Eynez and GLI-Tuncbilek) and some TTK underground hard coal mines using different methods together with various factors responsible for it, so that appropriate measures are taken to prevent or minimize them.

This aim will be achieved through the following objectives:

- Providing the basic statistics of accident data related to the lignite and hard coal mines
- Study of quantitative risk analysis technique for coal mining industry
- Probability analysis of accidents with the previous accident data
- Quantitative and semi-quantitative analysis of the given data for establishment of a relation between accidents and the expected causes, using the three methods - regression (Ordinary Least Square and improved), neural networks and fuzzy logic.
- Future accident estimation using regression and neural network techniques.

1.2. Organization of Thesis

Thesis is organized as follows. Basic definitions, reasons, identifications, theories and analyses in risk assessment are provided in Chapter 2. This chapter discusses the theory of hazards, mine risk management and accidents in Turkey and also explains the manner in which safety hazards are related to accidents.

Chapter 3 describes risk assessment techniques commonly used in mines which show the way of the study. This chapter includes; informal techniques, basic formal techniques and advanced formal techniques. Chapter 3 covers those factors which affect the performance of implementation of safety factors and previous risk assessment studies.

Chapter 4 reviews the literature on modelling methodologies which are Poisson distribution, regression analysis, parsimonious regression analysis, neural networks and fuzzy logic techniques. In the chapter previous studies on these modelling are also mentioned.

In Chapter 5, data collection and basic data analysis were done with accident data of ELI, GLI and TTK mines and a comparison was made between TKI and TTK mines according to simple statistics.

In Chapter 6, a risk analysis methodology is introduced; the probabilities, severities and frequencies are calculated considering accident reason, affected organs, shifts, seasons and age of the workers using probability, severity and frequency of the accidents.

Chapter 7 gives a brief description about statistical methodologies with pre-processing of data. In the chapter probability and severity components are discussed. The Poisson distribution, regression analysis, neural network and fuzzy logic are introduced in detail and modelling and analysis are performed.

In Chapter 8, Future accident estimation modelling technique is presented by using number of workers, number of working hours and total raw production of each mine.

Chapter 9 presents the results and discussions on the study, comparison of methodologies. Finally, Chapter 10 gives the conclusion and recommendations of the study.

CHAPTER 2

RISK AND RISK MANAGEMENT

2.1. Definitions

This section explains various definitions and technical terms covered in the thesis. Most of the definitions might be used differently under different contexts, so the most relevant one chosen. Definitions in quotes are adopted from (Ericson, 2005) unless otherwise stated.

1. **Accident** is defined as “Unexpected event that culminates in the death or injury of personnel, system loss, or damage to property, equipment, or the environment.” In the current context, accident will be frequently replaced by ‘incident’, and the definition also gets slightly modified to the following. An incident is called as an accident if it causes some loss of input expected from worker. Such a loss could be in the form of either loss of efficiency, or loss of hours or days.
2. **Hazard** is defined as “Any real or potential condition that can cause injury, illness, or death to personnel; damage to or loss of a system, equipment, or property; or damage to the environment.” For this study, it can be better defined as ‘any real or potential condition resulting in undesirable change in properties of accidents, such as increased workday loss per accident, or increased severity of accidents. For example, lack of awareness in workers about safety methods is an example of hazard. In this work, hazard has the same meaning as causes of accidents.

3. **Failure Rate** is defined as the reciprocal of amount of time required for 63.2% probability of mishap. The derivation of this will be discussed under the subject of hazard theory.
4. **Risk** is defined as "measure of expected loss presented by a potential event, such as a financial failure event, a schedule failure event, or a mishap event. In system safety it refers to mishap risk, where risk = probability * intensity of damage". According to this definition, risk is maximum when probability of occurrence of the event is 100%, which counters the definition provided by Singhal and Malhotra (2000), as a measure of uncertainty of event. This definition may be considered analogous to 'expected outcome', as defined by (Heizer, Rajashekhar and Render, 2009) in terms of operations management in the same manner. In other words, in the current context, risk is defined as a measure of expected undesirable effect of a hazard.
5. **Risk Assessment** is defined as "Process of determining the risk presented by the identified hazards." In current context, this description can be extended to a complete package of risk determination, risk treatment, risk monitoring & review, and risk related communication & consultation. Thus, it contains all steps of risk management after hazards have been identified.
6. **Risk Management** is defined as a combination of processes of identification of risky factors, assessment of risk associated with them and continual repetition of this process.
7. **System.** This is defined as "Composite, at any level of complexity, of personnel, procedures, materials, tools, equipment, facilities, and software. The elements of this composite entity are used together in the intended operational or support environment to perform a given task or achieve a specific purpose, support, or mission requirement." For instance, for a mine, the system might consists of persons like miners, supervisors, management, plus processes like movement of material, extraction, loading/unloading, communication et cetra.

2.2. Hazard Theory

Risk assessment of activities is a fairly subjective research area, and so, no ideal technique can be defined for it. Ericson (2005) proposed twenty two techniques of hazard analysis. Iannacchione, Brady and Varley (2008) provided for three broad types of classification for such techniques.

The definition of risk can be further extended with the help of a decision tree diagram, given in Figure 1. The first factor is task output, which can be divided in two or three parts. For instance, whether it is worth the efforts to start extraction operations in a coal mine, considering the fact that some mishaps might claim lives of workers or not. Risky operations should be initiated only if they are found profitable enough to cover up for that possibility of danger. This decision is beyond the scope of this research, because it is decided that the work is to be executed. Only thing which can be assumed is that work can have hazards only if some minimum return criteria are met. Second factor is hazard probability. Out of the various cases of return, the one with higher returns might be considered for relatively higher hazard probability, but the one with the lowest returns out of accepted (say average, if high & average were selected) will have a stricter selection criteria. Next factor is severity, which comes to picture if the combined effects of two factors neither approves nor rejects the situation. Since output is directly related to the three factors (frequency, probability and severity), risk is analogous to a product of output which is probability and severity. However where output is not considered, risk can simply be quantified as risk = probability * severity of accident.

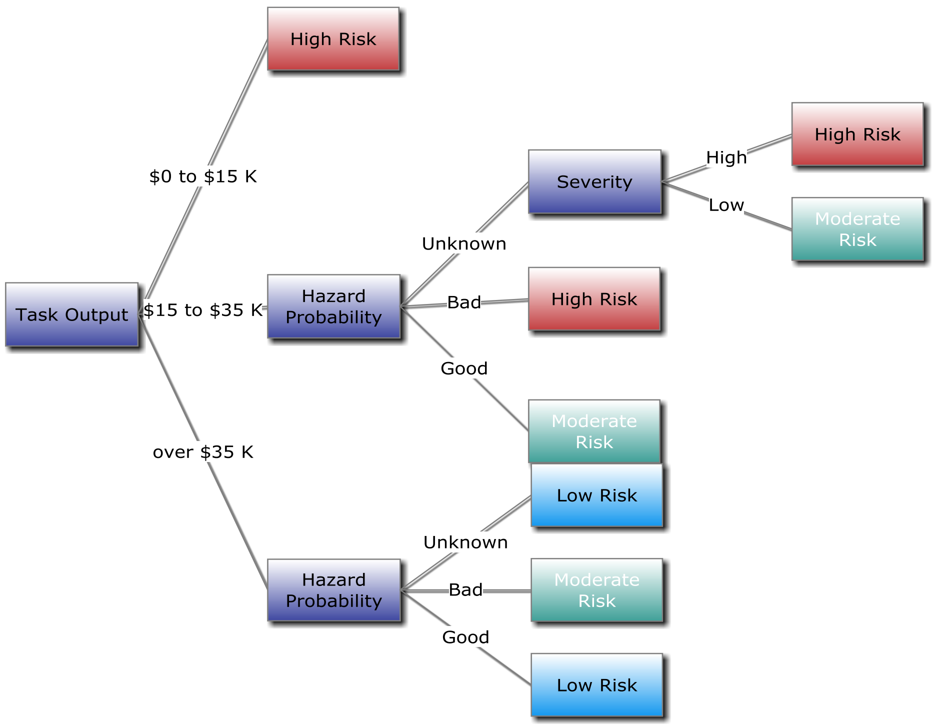


Figure 1. Decision tree diagram for risk (Yuan, 1995)

To reduce system risk, it is necessary to understand the root cause of such risks. According to definitions, risks are the expected outcomes of hazards – directly related to severity and probability. Thus, there are only two ways of reducing risks – (i) Reduction of severity of accidents, and (ii) Reduction of probability of accidents. In this work, the qualitative analysis deals with general methods which eliminate both (i) and (ii), both of whom require removal or reduction of hazards.

Since risk is a product of probability and severity, it is the one which governs the transformation of hazard to mishap. Risks originate from hazardous component, as will be discussed later in this chapter. Figure 2 shows transition of hazards to mishaps.

According to Figure 2, if the odds of risk factors associated with the system are in favour of mishaps, hazard components create mishaps. In other words, hazard is an 'initial and continuous' state of system, with some finite non-zero probability of transition to mishaps, which is an instantaneous event. The transition is immediate, though two things are not sure (i) Chances of occurrence (ii) Time of occurrence.

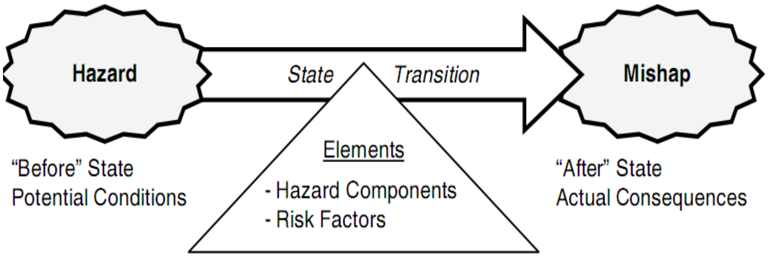


Figure 2. Transition from Hazard to Mishap (Ericson, 2005)

2.3. Mine Risk Management

This part covers the importance of risk management techniques, current scenario of mine industry and methods of identification. It also discusses the classification of hazard factors to internal and external.

Occupational Safety and Health Administration (OSHA, 2002) provides information on common hazards in industries, and the methods of controlling them. It also discusses on an analysis technique called job hazard analysis, according to which focusing on job task is a method of hazard identification. Thus, it attempts to link each worker to its tools, allocated task and the working environment. For instance, if a metal forming industry is considered, the method would focus on tasks like welding, cutting, shaping and drilling. For a driller, hazardous element could be drilling machine or the work piece itself, but for a cutting workman, hazardous element could be sharp edges of the tool itself.

OSHA (2002) also discusses the relevance and application of job hazard analysis. It could be summarized as (i) high injury rate jobs – ones where probability of hazards is high (ii) new jobs, or jobs with no previous records of accidents – high safety risk is expected from such jobs because unlike jobs with mishap history, new jobs are not prepared with safety measures (iii) jobs where process and procedures have changed very recently – for the same reason (iv) severe injury jobs – jobs with even low mishap probabilities require job hazard analysis, if mishaps are expected to result in very severe injuries because risk is a product of probability and severity (v) jobs where mishaps may result in severe financial damage (vi) Complex jobs – even if the work involves very small risk, it is possible that the employees make mistakes while taking instructions, such that in spite of being harmless, the accidents interfere with the normal working of the process.

OSHA (2002), further lists and describes several common hazards expected in production sector, as shown in Table 1.

Underground coal mining is an industry with the maximum hazard rate. For instance, a chemical hazard may result in fires (Stellman, 1998) and even explosion (Fesak, 1985). Stellman (1998) lists many other hazards associated with underground coal mining industry, such as extreme temperatures, poor visibility, weather and even ionization hazards. Similarly, there are hazards from mechanical sources (falling, transportation etc.). Thus, it can be said that coal mining industry involves most of the hazards listed in Table 1. This identification will be further used for quantitative analysis for establishment of relationship between reasons of accidents (hazards) and accidents. However, only a few factors will be chosen from the list for this analysis.

Table 1. Types of common job hazards and their descriptions (OSHA - 2002)

Chemical (Toxic)	A chemical that exposes the worker by absorption through the skin, inhalation, or through the blood stream, resulting in illness, disease, or death. The level (amount) of chemical exposure is critical in determining hazardous effects.
Chemical (Flammable)	A chemical that results in combustion when exposed to a heat ignition source. Typically, it is considered more flammable when it has lower flash point and boiling point.
Chemical (Corrosive)	A chemical that causes damage when it comes into contact with skin, metal, or other materials. Chemicals with extreme pH values (acids/bases) are common corrosives
Explosion (Chemical Reaction)	Self-explanatory.
Explosion (Over Pressurization)	Sudden and violent release of large amounts of gas (or energy) due to pressure difference between components and surroundings. Some examples are - rupture in a boiler / compressed gas cylinder.
Electrical (Shock/ Short Circuit)	Contact with exposed conductors or an incorrectly / inadvertently grounded device. For example, when a metal beam/ladder comes into contact with power lines. Even 60Hz alternating current (common house current) is dangerous enough because it can stop the heart.
Electrical (Fire)	Use of electrical power, resulting in electrical overheating / arcing to the point of combustion, ignition of flammables or electrical component damage.

Table 1 (continued)

Electrical (Static/ESD)	The movement or rubbing of wool, nylon other synthetic fibers against even flowing liquids can generate oppositely charged ions, resulting in static electricity. This ionized material surface may discharge (spark) to the ground, which may result in ignition/combustion of flammables, damage to electronics, or even damage to body’s nervous system.
Electrical (Loss of Power)	Failure of safety-critical equipment itself, due to loss of power.
Ergonomics (Strain)	Damage of tissues due to repetitive motion or overexertion (strains and sprains).
Ergonomics (Human Error)	An error-provocative system design, procedure, or equipment. (For example, a switch going up to turn on/off something).
Excavation (Collapse)	Soil collapse in an excavation or trench as a result of improper / inadequate shoring. Soil type is critical in determining the hazard.
Fall (Slip, Trip)	Conditions that result in falls (or slips) from heights or irregular walking surfaces (such as slippery floors, uneven walking floors, exposed ledges, poor housekeeping etc.)
Fire/Heat	Temperatures that can cause skin burns or damage to other organs. Heat source, fuel, and oxygen are required
Mechanical/ Vibration	Damage to nerve endings due to vibration, or safety-critical failure because of material.
Mechanical Failure	Self-explanatory; typically occurs when device exceeds design capacity, or is not maintained properly.
Mechanical	Skin, muscle, or any other body part exposed to cutting, tearing, crushing, caught-between, shearing items or equipment.

Table 1 (continued)

Noise	High noise levels (>85 dBA 8 hr TWA) resulting in hearing damage and/or inability to communicate safety-critical information.
Radiation (Ionizing)	Alpha, Beta, Gamma or X rays, neutral particles that cause injury (such as tissue damage) by ionization of cellular components.
Radiation (Non-Ionizing)	Ultraviolet, visible light, infrared or microwaves that cause tissue injury by thermal / photochemical means.
Struck By (Mass Acceleration)	Accelerated mass striking the body, causing injury or death. (For example, falling stones and projectiles.)
Struck Against	Injury caused to a body part, as a result of coming into contact with a surface on which action was initiated/performed by the person. (An example is when a screwdriver slips during the work.)
Temperature Extreme (Heat/Cold)	High temperatures result in heat stress or exhaustion, and low temperatures cause metabolic slow down such as hypothermia.
Visibility	Lack of lighting or obstruction in vision, resulting in an error or damage.
Weather Phenomena (Snow/Rain/Wind/Ice)	Self-explanatory.

As per the types of jobs suggested by OSHA (2002) for job hazard analysis, underground coal mining job falls in high injury rate, high mishap severity and high complexity work. Also, there are more works stating that underground mining industry is riskier than other similar industries, and even to the extent that workers in the business should be paid a premium for taking up the risk. (Epp et.al., 1977).

Apart from OSHA (2002), Stellman (1998) and Fesak et.al. (1985), Iannacchione et. al. (2008) also throws light on the importance of risk management, and various factors affecting the risks in mining sector which also states various forms of damages due to lack of risk management besides these Güyaguler (2000) noted that there are two major categories of cost resulting from accidents, usually referred as direct and indirect costs.

- mismanagement problems due to workers getting killed by accidents
- loss of work and delays in critical projects
- compensations and medical expenses
- damage to properties and infrastructure

For quantitative analysis, data segmentation and consolidation is required to ensure that the size of database, in terms of number of variables is not too large to handle. One way to segment it is on the basis of riskiness associated with the reason, based on preliminary analysis. OSHA (2002) listed a number of possible hazards in production type industries, but did not cover that much on their relative significance or riskiness. However, work done by Iannacchione et.al. (2008) discusses the types of hazards which were responsible for fatal mishaps on multiple occasions in US Minerals Industry.

Table 2, adopted from Iannacchione et.al. (2008), lists those hazard types, along with the number of death accidents associated with them. Such frequent and severe hazards might be considered as relatively more critical for risk analysis.

Table 2.Hazards associated with multiple fatalities in the US Minerals Industry (Iannacchione et.al. 2008)

Hazard Type	**Events**	**Fatalities**
Strata Instabilities (Struck by/against)	8	21
Explosions (Chemical)	4	33
Powered Haulage (Electrical)	2	4
Fire (fire/heat)	1	2
Equipment Failure (Mechanical)	1	2
Heat Strain (Ergonomic)	1	2
Slip or Fall of Person (Fall)	1	3

It can be seen that the data covers several types of job hazards given in Table 1 as well, but there are two differences. First, data grouping has been done on the basis of broad classification of hazard type. For instance, Strata instabilities can be considered as that of the following type: struck by, struck against. Similarly, power haulage can be put as ‘Electrical’. This classification is appropriate for this situation, if it is assumed that the dataset is exclusive – that is, the table does not contain more observation than the ones given. Second, it consists of only selected types of common hazards.

So far, some of the factors were identified, combined and grouped as well, but there is another basis which requires segregation – level of control. On other words, analysis of only those hazards is significant which can be controlled or eliminated. For example, rain is an initiating mechanism for hazards in a cement industry. Indirect methods can be used to eliminate the threat mechanism, but rain itself cannot be eliminated. In contrast, fire hazard can be eliminated by changing, say, working conditions.

2.3.1 External and Internal Factors

The report of NIOSH’s review committee (2007), defines external factors as those which cannot be controlled because either it is impossible to do so, or are beyond the scope of the committee’s research, but still affect the results. In contrast, internal factors are those which are supposed to be analysed, and on the basis of them, some methods might be recommended to modify or eliminate the effects of such factors. Similar definition applies for this work as well, and the factors will be dealt with accordingly.

This completes the classification of hazards on two grounds – physical nature and level of control. However, there could be another basis of classification – that based on the type of management levels. Frankel, Hommel and Rudolf (2005) explained;

- **Operational Risk** – This can be defined as a type of risk where the mishap may result in damage to process. For example: workday loss because of injury to worker, or inefficiency due to malfunctioning of some machine.
- **Financial Risk** – A type of risk in which loss is more of financial in nature. For example, damage to machinery because of initiation of fire.
- **Strategic Risk** – This type of risk affects working at higher levels. For example, if a report gets damaged due to loss of electricity.

The main focus will be on operational risks, since it is much more concrete and analysable using quantitative methods.

2.4. Risk Assessment

After the identification of risk factors, the next task is that of risk analysis, followed by risk evaluation, risk treatment, monitoring and review, and communication & consultation. All these factors constitute risk assessment, which is a part of risk management. Figure 3 shows the seven basic steps of risk assessment – (i) establishment of context, (ii) identification of risks, (iii) risk analysis, (iv) risk evaluation (v) risk treatment (vi) monitoring and review (vii) communication and consultation (BCI, 2007)

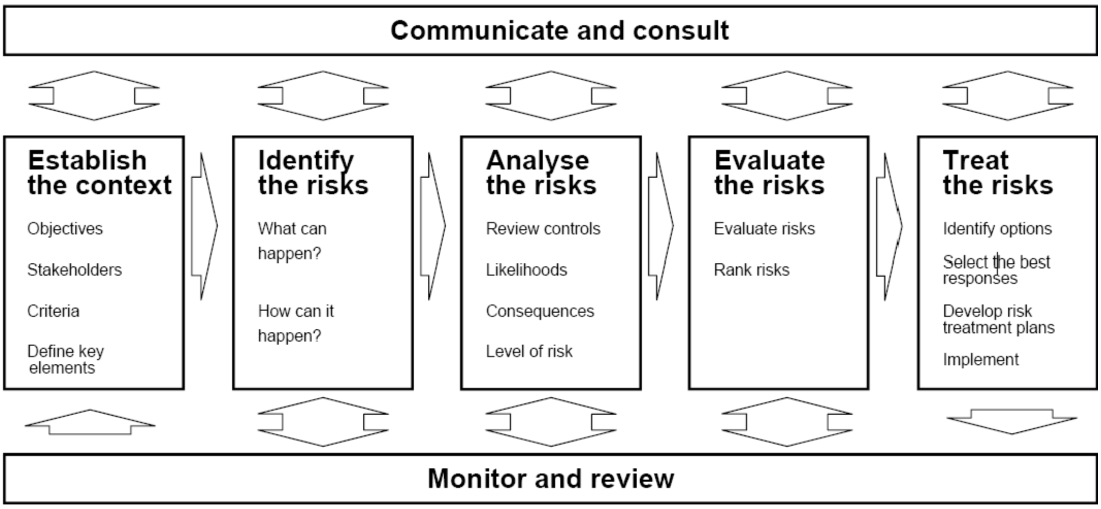


Figure 3. Seven basics steps of risk assessment (BCI, 2007)

Out of these steps, context has already been provided, so the remaining steps will be discussed here. The goal of risk assessment cycle is to minimize the risk to the level as low as possible.

2.4.1 Identifying the Risk

First step is risk identification – which means identification of hazards with high probability and severity. It has two parts:

- Descriptive – This method deals with risk identification through a comprehensive analysis of the entire system. This is done by the processes of (i) checking alignment between objectives at each level (ii) checking integration of levels against the ones immediately above/below them (iii) Measurement of partial impact of that particular level on overall system, and the effect on all stakeholders taken together. This is more suited for a system which is a part of a larger organization.
- Creative – In this method, the system is broken to several smaller units, which are then analysed separately. These units, known as key elements, ensure that all issues related to risk identification are properly covered.

Both methods are important for risk identification, and play their role simultaneously to ensure that the final result is optimum. Since risk identification is a complex process, efforts need to be made to simplify it. This simplification can be achieved either through checklist construction, or through a brain-storming session. The former is effective if all possible sources of risks (hazards) related to the system are known in advance. It assumes that the behaviour of the system is known, and is expected to remain as per that throughout. It has two disadvantages because of that assumption. First, system may not be known at all. Second, it does not consider the case of system behaviour going beyond the expectation. Second approach, brain-storming is usually preferred because it uses the power of creativity to come up with new and emerging sources of risk. It works on several levels, such as structured workshops, interviews, surveys etc. However, the downside is that it is unable to make very effective use of prior knowledge, because the method functions when all sources of prior knowledge are excluded from the session.

2.4.2 Risk Analysis

Once risk sources are identified, they are required to be analysed for quantification and attainment of relative picture of risks. Since severity and probability completely define a risk, the goal of analysis is to estimate the values of these two quantities associated with risks.

The method appears straight forward if it is assumed that it is possible to estimate both quantities, but there are several reasons why complications may still arise, and how they are required to be dealt with:

- If analysis is done by a group of individuals, their estimated values should be considered as a dataset for data analysis problem, from which some measure of central tendency (such as mean, median or mode) would provide the best set of values for the two. In case some non-central estimate is required (such as, worst case analysis), extreme values might be taken. For instance, if some floor of a room is left wet, it is the person working in that room who would be affected, and it is his experience which matters the most

- If analysis involves a complex risk, it might often be broken down to multiple risks. For example, if an electrician working on open wire slips off and dies, a new hazard might arise – open wires inviting new accidents. However, the two risks are inter-related and require a relatively complex analysis for value estimates

2.4.3. Risk Evaluation

This is the third step, and might be considered as a post-processing one for risk analysis. This is required for data adjustments and conversions, and involves steps like (i) re-processing of unexpected/incomplete/extreme data, (ii) compilation of data - continuous removal of risk factors rated too low (till the number of factors left is manageable for further analysis), and grouping of data, if required (iii) Adjustment of data according to the required tools, if any.

2.4.4 Risk Treatment

Once risks are identified, analysed and evaluated, they are finally required to be treated, meaning some suitable remedial action is suggested for them. For instance, if it is found that an immediate action of increasing temperature results in gas accumulation in boiler, one of the risk treatment measures suggested for the problem is to reduce the maximum allowed temperature for treating the risk.

2.4.5 Monitoring and Review

Risk assessment steps are complete at this point. However, to keep the system working at its full potential, the previous steps should be continuously monitored to ensure they are working as expected, and reviewed to take care of possibility of changing environment with time. For instance, one such change could be discovery of a more effective method of analysis, implementing which could result in enhanced performance. This means the system might not be optimized even if all five steps are working properly, and hence, monitoring & review step is required.

2.4.6 Communication and Consultation

Communication and consultation is also a parallel and continuous process like monitoring and review. Since risk management does not affect everyone directly in the same manner, effective communication is required to make sure that everyone is aware of hazards involved with the system, and the types of preventive or remedial measures to be taken to avoid it. Also, it is required to ensure that the safety needs of all stakeholders are addressed during risk identification step of risk assessment.

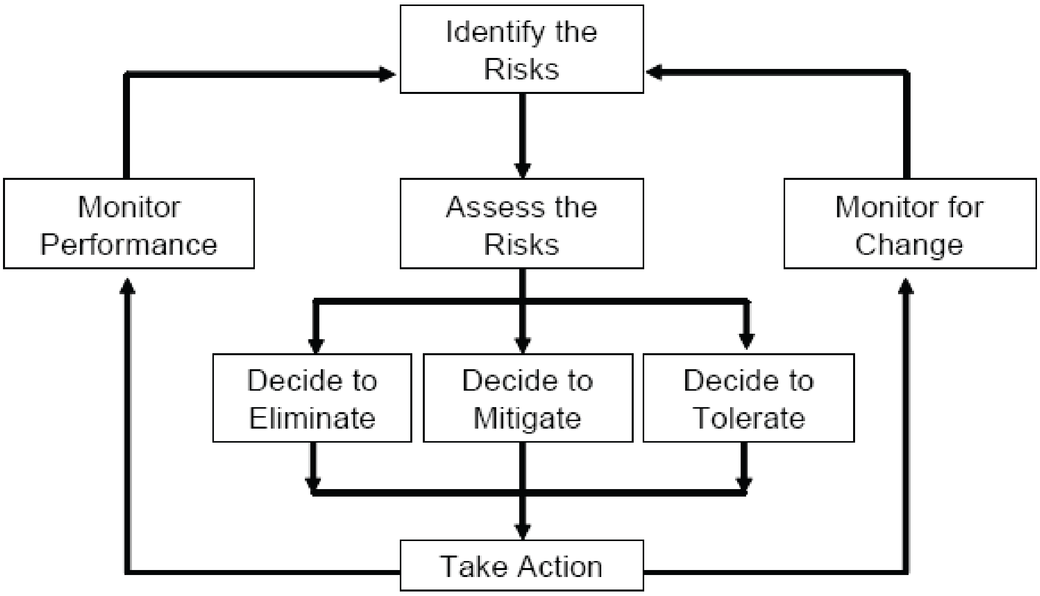


Figure 4. Risk Management Chart, (Iannacchione et.al., 2008)

2.5. Accidents in Turkey

According to statistics of labour, accidents in coal mining has a high trend comparing with the other sectors. In the last 5 years; 30 154 accidents happened in this sector. In Turkey; this number has the 8% of total accidents. It is expected that this rate is even higher considering lack of data collection and large number of unregistered workers.

Table 3, 4 and 5 presents basic data of coal mining, metallurgy, construction, textile and other industries in 2011, 2010 and 2009. (Social Security Institute, Yearly Statistics). Corresponding distributions are given in Figures 5, 6, 7

Table 3. Accident Statistics of 2011 (Social Security Institute Yearly Statistics)

Sector	Number of Workers	Number of Accidents	Number of Deaths
Coal Mining	51662	9217	55
Metal Works	515932	12540	90
Construction	1630851	7749	570
Textile	392550	3239	22
Other Industries	8439944	36482	973
Total	11 030 939	69 227	1710

Table 4. Accident Statistics of 2010 (Social Security Institute Yearly Statistics)

Sector	Number of Workers	Number of Accidents	Number of Deaths
Coal Mining	50143	8150	86
Metal Works	468665	11539	67
Construction	1450211	6437	475
Textile	356477	3474	16
Other Industries	7705314	33303	800
Total	10 030 810	62 903	1444

Table 5. Accident Statistics of 2009 (Social Security Yearly Statistics)

Sector	**Number of Workers**	**Number of Accidents**	**Number of Deaths**
Coal Mining	51975	8193	3
Metal Works	442865	12133	13
Construction	1227698	6867	156
Textile	331438	3771	12
Other Industries	6976226	33352	987
Total	9 030 202	64 316	1171

As it is seen from the tables even though the number of workers in coal mining sector is relatively lower than the other sectors, the number of accidents is quite higher.

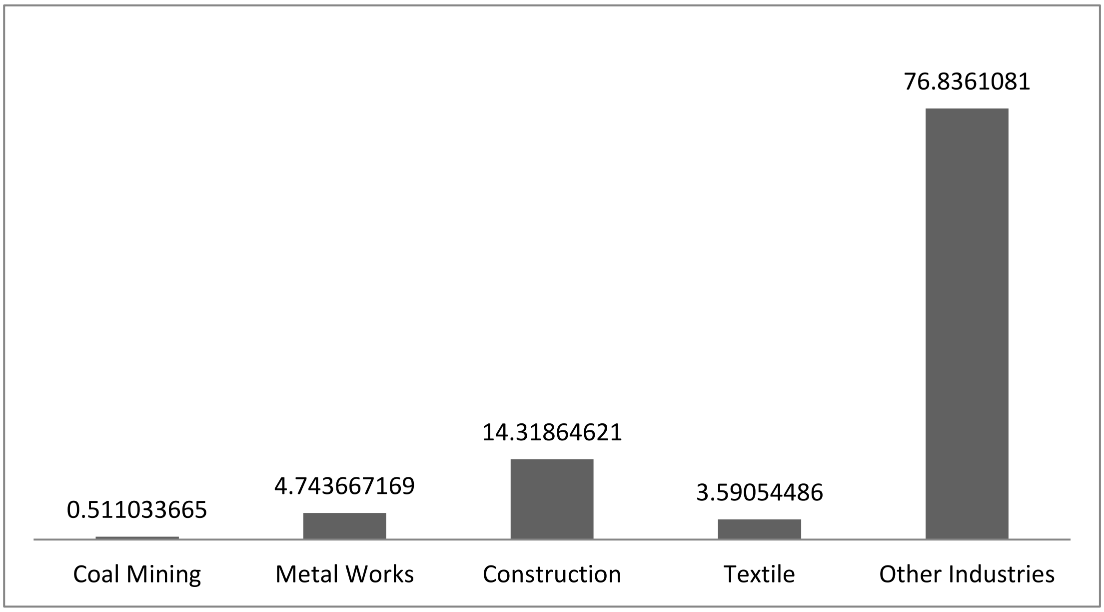


Figure 5. Distribution of percentages of workers

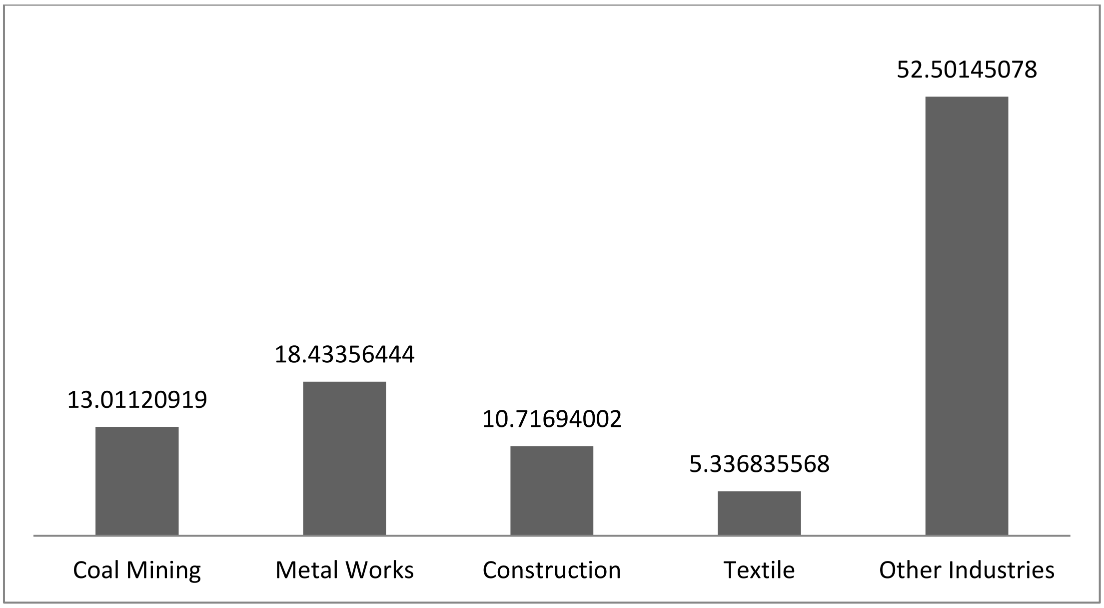


Figure 6. Distribution of percentages of accidents

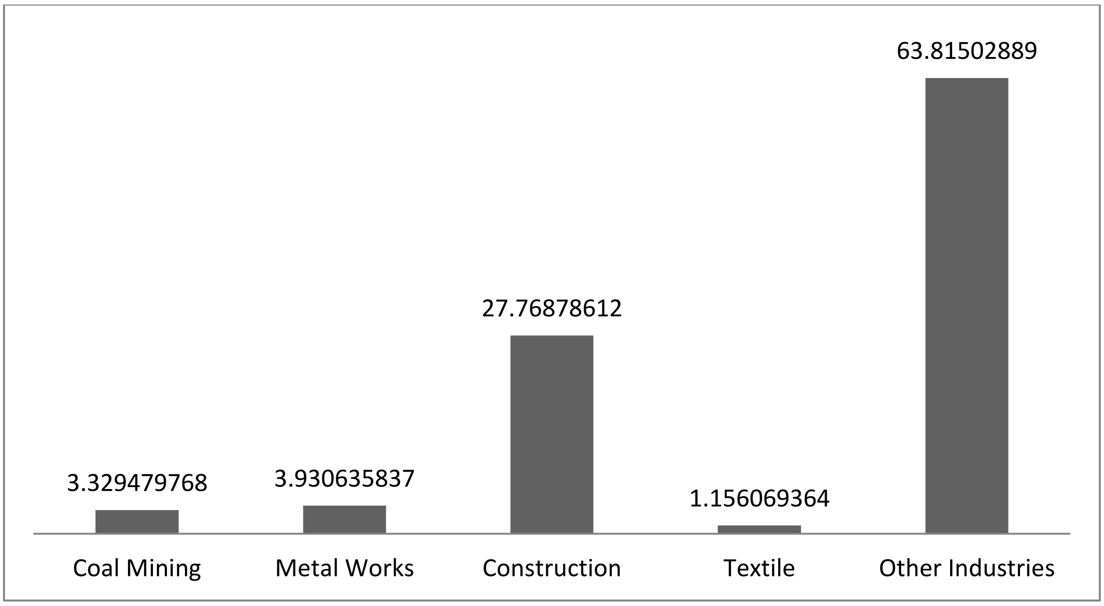


Figure 7. Distribution of percentages of deaths

When comparison comes to the number of deaths it is quite similar to metal industry while the number of workers quite higher than coal mining sector. Surface mining activities which are also less hazardous are included in the figures.

These figures show the accident rate and risk level of coal mining industry.

Table 6. Comparison of sectors for accidents and deaths per worker according to acerage numbers of 2009, 2010 and 2011

Sector	Accidents per worker	Deaths per worker
Coal Mining	0.166211471	0.000936403
Metal Works	0.025368101	0.000119092
Construction	0.004886093	0.000278734
Textile	0.00970323	4.62764E-05
Other Industries	0.004460657	0.00011937

Sarı and Karpuz (2001) studied about international comparison of Turkish coal mining industry safety performance. According to the study, Turkey has the highest fatality accident rates over other countries. The fatality rate in Turkey is 130 times more than the one in Australia. On the other hand, Turkish lignite coal data presents better rates.

CHAPTER 3

RISK ASSESSMENT TECHNIQUES IN MINES

There are several risk assessment techniques, but Iannacchione et.al.(2008) describe a broader classification for such techniques based on the levels of complexity involved. Going from simpler to sophisticated, they are (i) Informal Risk Assessment Techniques, (ii) Basic Formal Risk Assessment Techniques and (iii) Advanced Formal Risk Assessment Techniques.

3.1. Informal Risk Assessment Techniques

Such techniques just require worker to look for possible hazards, determine the risk carried by it and take suitable action to minimize it. A few examples are as follows:

- SLAM (Stop-Look-Analyse-Manage) – it requires workers to halt the work, and examine and analyse the process, and ask for risk treatment if required
- Take-Two for Safety – requires the worker to think for two minutes before resuming work
- Five-Point Safety System - requires that workers themselves take up the responsibility for processes within their scope
- Take Time, Take Charge – requires workers to them to respond to hazards, if required.

3.2. Basic-Formal Risk Assessment Techniques

These techniques require adherence to a set of rules and procedures to be followed prior to performing high-risk activities. So, they also include a requirement documentation and subsequent monitoring of activities. Being formal, such techniques are less simple than informal ones. However, they do not use sophisticated methods.

An example of most commonly used technique of this type is Job Safety Analysis (JSA). This is very similar to Job Hazard Analysis (JHA) by OSHA (2002), and is also used for processes prone to high risks, but JSA uses Standard Operating Procedures (SOP) for documentation. Other examples of Basic Formal Risk Assessment Techniques are CTA (Critical Task Analysis) and JHB (Job Hazard Breakdown).

3.3 Advanced-formal risk assessment techniques

They are used when Basic Formal Techniques do not suffice for analysis purpose. These techniques require structured approach with document, but in addition, they also incorporate applications of some risk analysis tools, such as Workplace Risk Assessment and Control (WRAC), Preliminary Hazard Analysis (PHA), Failure Modes, Effects and Analysis (FMEA), Fault/Logic Tree Analysis (FTA / LTA), Hazard and Operability Studies (HAZOP), Bow-Tie, as will be discussed next. If one analysis does not suffice for analysis purpose, a combination of techniques might be applied. Such a combination should be carefully selected because each technique requires more time and effort than Informal and Basic Techniques.

Workplace Risk Assessment and Control (WRAC)

This techniques work by breaking down the system by parts of mine involving different types of processes, and then ranking them on the basis of risk rating. Risk was earlier defined as a product of Likelihood (probability) and Consequence (severity), but some other risk function (r = f (l,c)) can be used. The procedure goes as follows:

i. Divide the system to parts, note down the potential unwanted event related to that particular part, and lay them down in the form of matrix.
ii. Calculate consequence and likelihood columns of the matrix.

iii. Compute r = f (l, c), and rank the parts on the basis of values of r.

For the above procedure, consequence may be computed based on the cost of equipment at site, and the effective cost of loss of workday, death insurance, damage to goodwill etc. wherever required. Some typical methods of calculating the likelihood of mishap are (i) Previous History (ii) Average Estimate (iii) Worst Case Estimate.

Preliminary Hazard Analysis (PHA)

This method is very similar to that of WRAC because it ranks the potential hazardous events, but there are two differences (i) it focuses on all events of the system, instead of parts of the system, and (ii) It only performs analysis on the most likely consequence, instead of average consequence. The format includes total exposure which is not exists in WRAC.

Failure Modes, Effects and Analysis (FMEA)

This method focuses on failure modes, its effect on the current item, and the system as a whole; instead of just events and affected item. Also, the final result is a control measure in response to the level of criticality (risk). In other words, the analysis of a particular item is absolute and exclusive from other items. The advantage of this method is that the extent to which risk will be treated depends on criticality more than on rank. However, on the downside, it does not consider the priorities. Table 7 shows how FMEA is implemented.

Table 7. FMEA implementation (Ericsson, 2005)

<table>
<tr><td rowspan="3">Item</td><td rowspan="3">Failure Mode</td><td colspan="2">Effects on</td><td rowspan="2">Likelihood</td><td rowspan="2">Consequence</td><td rowspan="2">Risk</td><td rowspan="3">Control</td></tr>
<tr><td rowspan="2">Other items</td><td rowspan="2">System</td></tr>
<tr><td>(L)</td><td>(C)</td><td>(LxC)</td></tr>
</table>

The algorithm can be defined as follows:

i. List the types of failure modes possible, and their effects on the system.
ii. Calculate risk using method described in algorithm for WRAC.

Fault/Logic Tree Analysis (FTA/LTA)

This type of analysis is useful for situations involving events with multiple possible consequences, and the effect of probability of all of them is required to be considered. FTA is similar to LTA except that the former focuses on finding the most unwanted one (also called ‘fault’), whereas LTA emphasis on finding the most wanted one.

Fault Tree model is preferred, but LTA may also use the logic-gates model, where each gate represents some binary operation, and the result is a set of hazards with risks worth consideration. However, both the methods require probabilities of events to be known beforehand.

Hazard and Operability Studies (HAZOP)

HAZOP is used in situations where small deviations in a process may lead to big hazards, and it is possible to analyse this beforehand to a reasonable extent. The main sector which uses this type of assessment is chemical industry. It involves performing of what-if analysis through an instrumentation process, and can be used for modelling and projection of various types of properties, such as fluid flow, temperature, pressure, concentration and so on.

Bow Tie Analysis

In this type of analysis, the entire problem is arranged in the form of a bow. Undesirable event, which is the result of other elements, is kept at the centre of the bow. Threats, their causes, and corresponding control measures occupy the left side of the tie bow (also known as prevention side). Consequences, recovery measures to deal with them, and the potential outcome take the right hand side of the bow (also known as recovery side). The entire bow signifies the following: Lack of preventive measures result in unwanted events, and those events result in consequences. The entire picture is demonstrated in Figure 8.

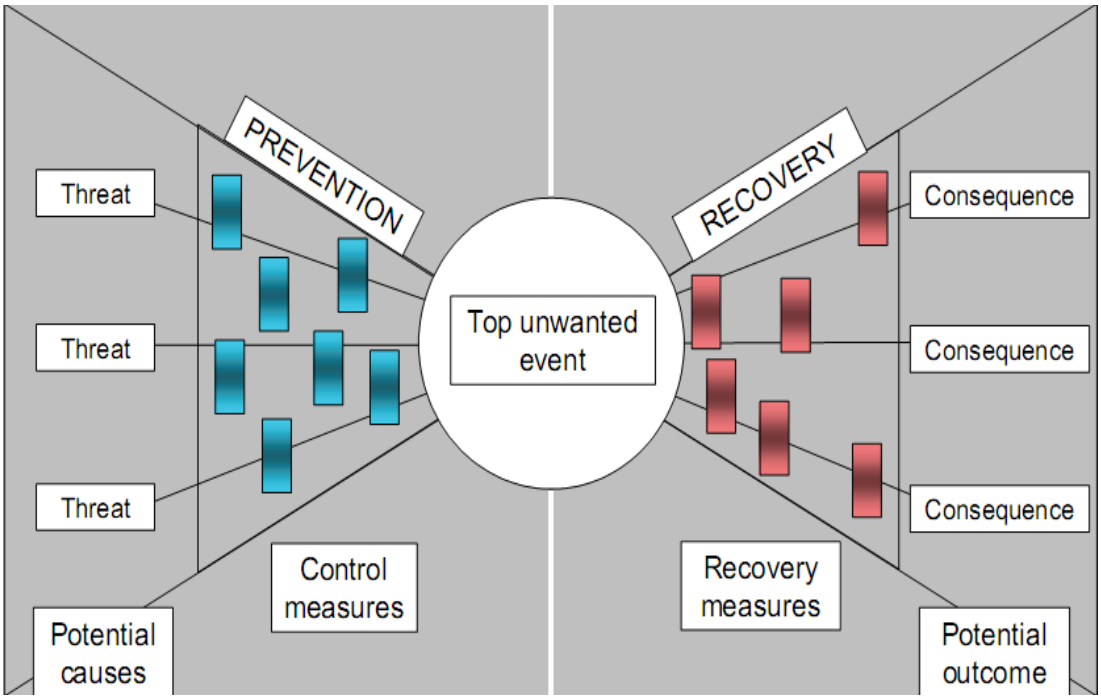


Figure 8. Bow Tie analysis method (Iannacchione et.al. 2008)

3.4 Factors Affecting Safety Performance in Mines

This section discusses those factors which affect the implementation of safety performance in mines. Sari (2002) discussed and elaborated on a few of them:

- Depth of mining field (deep/shallow) – it is possible that safety measures found to make significant improvement for underground mining may not work (or be required for surface mining)
- Concentration of employees – if the system's employee base is huge and complicated, it is possible that some of them are unable to utilize some standards due to lack of information

- Amount of automation - high automation might make it difficult to explain and implement safety standards. On the other hand, if larger part of process is by hand, workers get aware of safety practices by experience as well
- Size of mining establishment – similar to that of concentration of employees, if the size of establishment itself is huge, some safety measures may not be implemented due to information getting lost in communication.
- Mining method and type of coal used – if mining methods are too complicated for employees to understand, rate of error becomes high and thus, safety methods become ineffective.

Overall, coal mine safety is influenced by many factors, that is to say, coal production systems is a system composed of personnel, machinery, equipment and extremely complex space. (Yang, 2010) Table 8 shows the coal mine safety evaluation with effecting factors.

Table 8. Coal Mine Safety Evaluation Index System (Yang, 2010)

Overall Target	Sub-targets
Coal Mine Safety Evaluation System	Mine geological factors
	Mine disaster factors
	Mine hazard factors
	Environmental condition factor
	Production staff quality
	Factors of production equipment
	Management factors

3.5 Previous Risk Assessment Studies

In the past years many scientists conducted studies about hazard identification and occupational safety risk assessments in various sectors. Sarı (2002) made a risk assessment approach on underground coal mine safety analysis and he determined the risks in ELI-Eynez and GLI-Tuncbilek mines by statistical analysis of past accident/injury experience data. Sarı (2002) designed the methodology for aforementioned mines considering the data of 1996 and 2000 covering 5 year period accidents. In the study of Sarı (2002), the methodology developed according to following steps.

- Identification of accidents in the mine
- Evaluation of probability of accident occurrences
- Determination of magnitude of accidents by establishing possible consequences or severity.
- Compilation of probability and consequence (severity) under a risk formulation
- Establishment of risk levels based on severity and probability
- Setting an acceptable risk level
- Risk management and control methods

In his research first the accidents were identified and the magnitude of harmful effect was determined by finding the frequency of different type of accidents. Risk levels were determined and the risk classification matrix was proposed.

Table 9. Risk classification matrix (Sarı, 2002)

	Severity			
Probability	1 - 10	10 - 100	100 - 1000	1000 - 10000
0.75 - 1.00	Moderate	High	Very High	Severe
0.50 - 0.75	Low	Moderate	High	Very High
0.25 - 0.50	Very Low	Low	Moderate	High
0.00 - 0.25	Very Low	Very Low	Low	Moderate

Fine and Kinney, 1976 found a practical methodology for safety risk of naval sector. According to them; the risk imposed by some particular hazard can be taken as increasing (1) with the likelihood that the hazardous event will actually occur, (2) with exposure to that event, and (3) with possible consequences of that event. For risk calculations, numerical values are assigned to each of these three factors.

Marhavilas (2009) conducted a study about risk estimation in construction works considering Fine and Kinney (1976) methodology, according to Marhavilas (2009); Risk can be quantified and can be measured by a mathematical relation which is;

$$R = P.S.E \tag{3.1}$$

where:

R: the Risk

P: the Probability Index

S: the Severity Harm Index

F: the Exposure Index

Table 10, 11 and 12 give the description of undesirable event for probability, severity and frequency.

Table 10. Gradation of the Probability Index in association with the undesirable event (Marhavillas, 2009)

Probability Index (P)	Description of Undesirable Event
10	Unavoidable
9	Almost assured
8	Very Probable
7	Probable
6	Probability slightly greater than 50%
5	Probability 50%
4	Probability slightly less than 50%
3	Almost improbable
2	Very improbable
1	Improbable

Table 11. Gradation of the Severity of Harm Index in association with the undesirable event (Marhavillas, 2009)

Severity of Harm Index (S)	Description of Undesirable Event
10	Death
9	Permanent total inefficiency
8	Permanent serious inefficiency
7	Permanent slight inefficiency
6	Absence from the work >3 weeks, and return with health problems
5	Absence from the work >3 weeks, and return after full recovery
4	Absence from the work >3 days and <3 weeks, and return after full recovery
3	Absence from the work <3 days, and return after full recovery
2	Slight injuring without absence from the work, and with full recovery
1	No one human injury

Table 12. Gradation of the Frequency Index in association with the undesirable event (Marhavillas, 2009)

Exposure Index (F)	Description of Undesirable Event
10	Permanent presence of damage
9	Presence of damage every 30 sec
8	Presence of damage every 1 min
7	Presence of damage every 30 min
6	Presence of damage every 1 hr
5	Presence of damage every 8 hr
4	Presence of damage every 1 week
3	Presence of damage every 1 month
2	Presence of damage every 1 year
1	Presence of damage every 5 years

Table 13. Gradation of the Risk Value in association with the urgency level of required actions (Marhavillas, 2009)

Risk Value (R)	Urgency level of required actions
800-100	Immediate action
600 - 800	Action during 7 days
400 - 600	Action during 1month
200 - 400	Action during 1 year
<200	Immediate action is not necessary but it is required the event surveillance

The Probability Index is calculated for accidents of specific for the group by using the corresponding number of accidents and the equation is;

P = (Number of accidents / Total number of accidents) x 10 (3.2)

The severity index is estimated for the worst case of the specific accident using loss workday cases and comparing with the table. The Frequency Index (F) shows the number of accidents during a definite time period. In order to calculate the accidents' frequency (per day), data for 10 year time period (assumed 50 working weeks and each working week with 7 working days) in the relation:

$$F = \text{Number of accidents} / (50x7x10) \quad (3.3)$$

Radosavljević (2009) used the methodology which is originally found by Fine and Kinney (1976) using qualitative information combined with quantitive analysis.

According to Radosavljevic (2009), the sources of information for the subject analysis are;

- The experience of following the process of the coal processing functioning,
- Interviews with subjects who are either direct or indirect participants within the working process,
- Testing the collected data,
- The history of casualties connected with the process of coal processing
- Management expert meetings, (the available data),
- Notes from scientific gatherings and symposiums whose subject was the risks and safety in mining/the process of coal processing and
- Experts' assessments and suggestions connected with the problems of the risk analysis.

Besides these, many scientists used risk assessment approach in different sectors and measuring different risks apart from safety risks. For example Arda (2008) used risk assessment technique for analysing the risks of a textile plant. Bu-Quammaz (2007) used the methodology for analysing economical risks in international construction projects similarly Karadas (2007) made risk analysis in defence industry to use in the business development phase.

CHAPTER 4

LITERATURE ABOUT MODELLING METHODOLOGIES

4.1 Regression Analysis

Regression is using to find the equation of a single line in the given input parameters which fit the given data. (Lind, D. A. , Marchal, W.G. and Wathen, S.A. , 2005). This is the simplest method, since if all values of input parameters and corresponding output values are available in continuous form for the given dataset.

The regression equation can be written, in its simplest form, as follows:

$$Y = R * X + C \tag{4.1}$$

where

R = Corresponding regression coefficient for that variable – this is the quantity to be found from regression, which produces the best fit for the line.
X = Value corresponding to that point (or 0/1 for dummy variable for the corresponding state)
C = Value of constant term
Y = output values for regression

Since dataset has more than one input parameters, the regression form to be used will be ‘multivariate regression’. Accordingly, the quantity X will be a 2-dimensional matrix, with columns representing input quantities, and R will be a vertical matrix (now replaced by β). Equation can now be modified as follows:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}} \\ \mathbf{x}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}. \tag{4.2}$$

One of the best and most commonly used estimate method, suggested by Lind et.al. (2005) is least squares method. In this method, the line parameters are computed such that the sum of squares of distances between data-points and the line is minimized. In other words,

$(\delta/\delta x)\sum\{ \sum(X_j - X_{ji})^2 + (Y\text{-}Yi)^2\} = 0$

$(\delta/\delta x)\sum\{ \sum(X_j - X_{ji})^2 + (Y\text{-}Yi)^2\} = 0$

where i,j represent points and input variables respectively, and other symbols have usual meanings.

On solving, it comes out to the following:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} = \big(\tfrac{1}{n}\textstyle\sum \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}} \big)^{-1} \big(\tfrac{1}{n}\textstyle\sum \mathbf{x}_i y_i \big). \tag{4.3}$$

This result can easily be found using some engineering or data analysis tools. For example, if Matlab is used for programming, the function ‘regress’ accepts matrix forms (as given in equation 4.3) of input and output variables, to produce various statistical quantities, such as regression equation, correlation coefficients, errors and so on. However, as already stated, it is difficult to analyse the set of parameters as given in Table 14 because the set contains both continuous and discrete variables. Numbers related to such ‘discrete’ values are ordinal, and do not represent any relationship. For example, it will be wrong to state that shift two is twice of shift one, and so on. So, it is necessary that such quantities are converted to some other format which makes them look like continuous.

One such format was suggested by Hardy (1993), who explained the method of using dummy variables for regression analysis. As per the method, if variables are represented by a set of binary quantities, they can be made to behave like continuous variables. In the previous example, the behaviour of ‘shift’ can be made continuous, if three shifts are represented as given in Table 14, by variables V1, V2 and V3, such that V1 represents (S1,0,0) and so on. Thus, continuous variables V1, V2 and V3 will take the required shift coefficients for states S1, S2 and S3 respectively. For instance, coefficients of (V1, V2, V3) = (1, 3, 7) represent the values 1 if state = (S1, 0, 0) ,or simply state S1 = True. Similarly, value becomes 3 for if state = S2, and 7 if state = S3. Thus, each parameter generates the number of variables equal to the number of discrete values it may take.

Table 14. Representation of shift through three dummy variables (S1, S2, S3)

Shift	V 1	V 2	V3
S1	1	0	0
S2	0	1	0
S3	0	0	1

There are inbuilt functions to convert the variables to the format described in Table 14. For instance, function ‘dummyvar’ in matlab converts a discrete variable carrying values 1-n to dummy variables. However, unless states are already provided in natural numbers format (1, 2, 3), the mapping of states with natural numbers need to be done through implementation of logic through a code or manually.

As will be established through regression is easiest to design, implement and get processed by machine, but it is not used too often because it is found that some other methods produce better results.

4.2. Parsimonious Regression Analysis

Some recent researches suggest that ordinary least squares regression can be significantly improved if some parameters, which do not correlate well enough against the given outputs, are removed (Sonmez et.al. 2007). This can be verified by analysing whether a parameter acts as an ‘outlier’, as per some distribution which represents the data. Most commonly used distribution is a ‘normal’ distribution, which can be assumed if the data set has more than 30 points (Lind et. al. 2005). So, under that assumption, a majority of points should lie close to the mean value, which should be true for each independent parameter’s relationship against the dependent parameter. If y depends on (x1, x2, x3….), and the value of Pearson coefficient (measurer of the extent of fit for some xi against y on the basis of fit of x1,x2…xn against y) of some xi’s against y is higher than the pre-decided threshold value, the parameter xi could be acting as an outlier, and can be removed. Then, the regression for the remaining set is performed. This method is slower than one-step regression, but sometimes produces better results.

4.3. Neural Networks

This method was first discussed by Gurney and Gurney (1997). It works in a way similar to regression, but uses past available values, or ‘learning system’ for analysis, instead of one-time method like least squares. It initially creates a hidden layer of neurons, pre-decided and unalterable parameters. The structure becomes similar to Figure 9. The layers input and output contains the information provided in training set, whereas the hidden layer contains information on weights of regression coefficient, which get modified after each cycle.

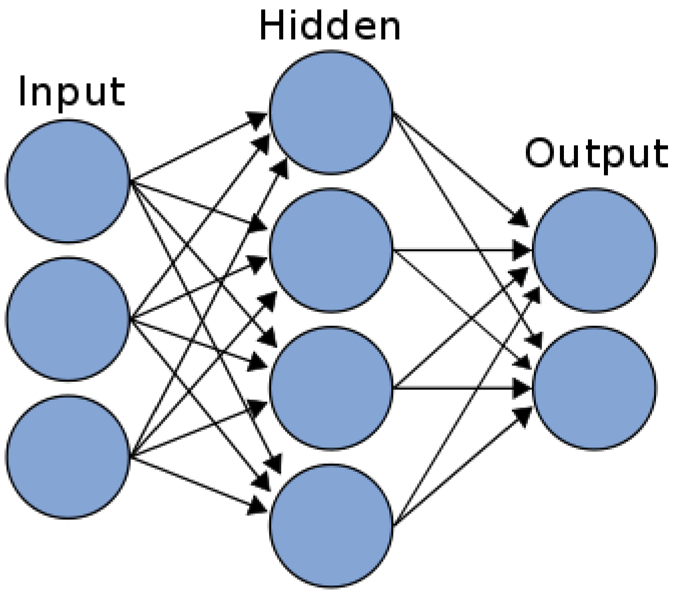


Figure 9. Demonstration of a typical Neural Network

Once weights are obtained through training, given system uses those weight values to get actual results from its database. For regression analysis type problem, a part of the given observation set itself acts as a training set and the remaining as a testing set. A particular percentage of points (10-20) is also used for validation of result quality. Matlab's Neural Network toolbox uses for analysis, by default allocates random 70% of the available data points for training. Out of the remaining, half (15% of total) of the points are chosen for data validation (validation set), and the remaining for testing the method (testing set).

The entire algorithm, as provided in the help sections and topics of Matlab's neural network method, goes as follows:

i. Prepare the inputs in the same way as those for regression.
ii. Select a random seed for process initiation, to ensure that the results do not vary with the experiments (this value cannot be changed by user).
iii. Prepare the network according to first (or nth) value of number of neurons.
iv. Get initial estimates of coefficient values, and calculate the error value I. Choose some learning value (a)

v. Modify all weights such $w_i = w_i - e*a*x_i$, and calculate new error values according to the new weights.
vi. Keep repeating till the desired learning level is achieved.
vii. Apply the network to the set of points allocated for testing.
viii. Validate the method efficiency by applying the network to the validation set.

The functions require data in the form of matrices of inputs and outputs on which analysis is required to be done. Such a conversion may be achieved through a set of software codes or by excel programming.

4.4. Fuzzy Logic

Zadeh (1965) has introduced fuzzy set theory as a mathematical useful tool for modelling uncertain (imprecise) and vague data in real situations. The essential assumption of fuzzy set is that many sets in real world do not have precisely defined bounds and each element has degree of belonging to some sets called as membership. (Song, 2005)

Fuzzy Logic method is frequently used to model the data when either the results are qualitative, or the number of variables is huge. The advantage of this method lies in the fact that it exploits the power of intuitions and common sense to come up with the expected results. On the other hand, it requires human intelligence and creativity for writing rules well-enough or else, the method may fail.

The basic idea behind the method is that related to flexible and uncertain reasoning, that every apparently binary system may take any values between 0 and 1. For instance, a two-value logic analysis might allocate the values 0 or 1, to say, some fluid when it is hot or cold respectively. In contrast the same liquid might take either of the following values under fuzzy logic – 0.5 for moderate, 0.25 for hot, 0.75 for cold, 0.9 for freezing, and 1 for minimum possible temperature. The user might define certain actions depending on the state. (Ross, 2009) This example is elaborated in Table 15.

Table 15. An example of fuzzy logic implementation for variable 'Temperature'

Binary Value	Definition	Condition	Action
0.25	T ~ 400	Hot	Start the Air Cooler
0.5	T ~ 298	Moderate	Do Nothing
0.75	T ~ 210	Cold	Start the Heater
0.9	T ~ 70	Freezing	Set the Heater Temperature to Maximum Value
1	T ~ 0	Absolute Freeze	Condition Impossible – Examine the measuring devices

From Table 15, it can be seen that some actions have been decided based on temperature values. Such a perspective enables decision making depending on whether the given value is closest to the given condition. For instance, if binary value corresponding to the temperature approaches 0.89, it might mean that the heater's maximum capacity is about to reach, and it needs to be shut down. This example involved some precise and continuous measurable quantity (temperature), but there could be systems having no such quantity. For example, satisfaction level of customers cannot be directly measured, but management might come up with the similar set of rules using its intuitions, skills and experience.

Similarly, for current situation, it can be said that the measurable quantity contains a combination of contribution of fuzzy values corresponding to values of its inputs. For instance, if the quantity is cumulative financial loss, it would be related to the sum of fuzzy values of inputs to cumulative financial loss, such as reason of accident or the organ of the injured affected by it. On contrary, the problem is related to measurement of such individual efforts based on the given data. Also, the result of the efforts (y) for a particular accident might have a purely random value, which is nowhere related to the values of inputs.

Fortunately, for huge number of data points, it can be safely assumed that an experiment set corresponding to one value of a variable is complete, and covers all possible combination of effects of other variables. For example, if the entire dataset available for second shift is considered, it can be said that the mean value of all such observations should be different from a similar set for third shift exclusively because of the difference in shift values. This difference in effects of outputs for different states can be measured in several ways – one of them is average value of experiment. This concept can be slightly modified for continuous variable by using class groups instead of discrete states, such as, average output values for weights of workers in the range 50-60. In other words, the weight contributions of all those observations whose weight values fall in the range 50-60 will carry this value. Similar values may be computed for other variables, such as reason, affected organ etcetera.

Once all values are calculated, a 'rules' set needs to be written corresponding to all possibilities. For example, what is the most probable effect on a worker aged 35-40 and weight 60-70 kg who is hit by a piece of scrap falling on his foot? The number of similar cases would be the product of number of values (or classes) each variable may take. For instance, if there are 2 classes for weight, 3 for age and 5 possible reasons of injury, there could be a maximum of 30 rules. After these rules are written, they have to be segmented to blocks such that fuzzy function can be applied on them. For each segmentation, there would be one function. If there are only 3 classes (say, output = high, medium or low), there will be one function corresponding to low, one to medium and one to high. Low function will be characterized by having low value when function's value is high and so on.

4.5. Previous Studies

Regression analysis is one of the basic modelling methodologies used in many studies. Mustakim, et al (2008) developed a traffic accident prediction model using multiple linear regression his study involves the identification of accident blackspot locations, establishment of general patterns of accident, analysis of the factor involved, site studies, and development of an accident prediction model using Multiple Linear Regression.

Sonmez (2004) used the model for conceptual cost estimation of building projects. He used parsimonious model which can be defined as a model that fits the data adequately without using any unnecessary parameters. He developed a backward elimination method in which all of the independent variables were considered in the initial regression model and variables that were not contributing to the model were eliminated one at a time.

Sonmez (2004) used also neural network in his study and compared both regression and neural network models.Comparing the models he used Mean Squared Error (MSE).

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\text{actual}_i - \text{predicted}_i)^2$$

Moghaddam,et al. (2010) used artificial neural network for predicting of accident severity in highways. He developed multilayer perceptron (MLP) that is a feed forward network in which information flow from input side and pass through the hidden layers to the output layer to produce outputs. Besides the study of Moghaddam (2010) neural network was used in many traffic safety modellings by Hashemi (1995), Chong (2004), Sando (2005).

Apart from accident severity estimation models, neural network gives opportunity to use in other elements. Karacan (2007) used the models for development and application of reservoir models and optimizing ventilation air requirements in development mining of coal seams. Ruilin (2010) used for prediction of coal and gas outbursts using Chinese coal mines statistics and Hong (2010) applied the model for gas warning system.

Yang (2010) used neural network for coal mine safety evaluation with V-fold cross-validation. According to Yang (2010) the neural network process is shown in Figure 10.

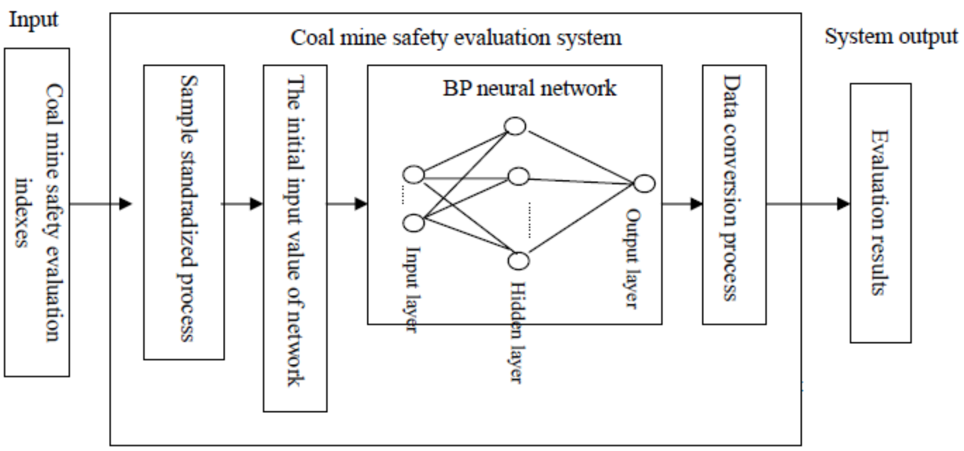


Figure 10. Flow chart of BP neural network training process. (Yang, 2010)

Fuzzy logic is also used in many prediction techniques. Gurcanli (2005) applied the technique in hazard assessment on construction sites. By this approach historical accident data in the industry are incorporated into the method. These inputs, subjective judgments of the experts and the current safety level of a construction site are combined by the utilization of fuzzy rule based system.

Sii and Wang (2004) used fuzzy logic for safety assessment in maritime engineering applications. According to Sii and Wang (2004) there are five steps in the fuzzy inference process.

- Step 1. Fuzzification of input variables - The first step is to transform the inputs into degrees of match with linguistic values via membership functions.
- Step 2. Application of fuzzy operations (AND or OR) in antecedents. This value will then be applied to the output function. The input to the fuzzy operator may have two or more membership values from fuzzified input variables. The output is a single truth value.

- Step 3. Implication from antecedent to consequent - The single truth value of a rule is determined by AND operator of the rule antecedents. With AND operator, rule evaluation then determines the smallest (minimum) rule antecedent, which is taken to be the truth value of the rule.
- Step 4. Aggregation of consequent across the rules - The output of each rule is combined into a single fuzzy set through the aggregation process.
- Step 5. Defuzzification - Finally, defuzzification process transforms the fuzzy results (i.e. a range of output values from the aggregation process) into a crisp output.

Mancini and Masi (2012) used fuzzy logic for environmental hazard assessments and Han (2005) developed fuzzy model for estimation of cost overrun risks in international projects.

CHAPTER 5

DATA COLLECTION AND BASIC STATISTICS

Data utilized in the study was provided from two underground lignite mines of Turkish Coal Enterprises which are GLI Tuncbilek Underground Lignite Mine and ELI Soma Underground Lignite Mine besides these ; Turkish Hard Coal Establishment (Zonguldak) coal mines accident data was collected and the comparison was made between three different areas.

Turkish Coal Enterprise's data is covered 14 year period of time from the years of 1997 to 2011 and TTK data covers 4 year period from 2008 to 2011.

5.1 ELI Soma-Eynez Underground Lignite Data Analysis

ELI Soma – Eynez coal mine is one of the oldest and biggest lignite mine which is located in Manisa.

Total 1033 accident data of January 1997 – December 2011 was used in the studies. The data covers, workday loses, affected organs, reason of the accident, shift, date of accident (season) and age of the victim.

The basic statistical analyses results presented in Figures 11 to 15.

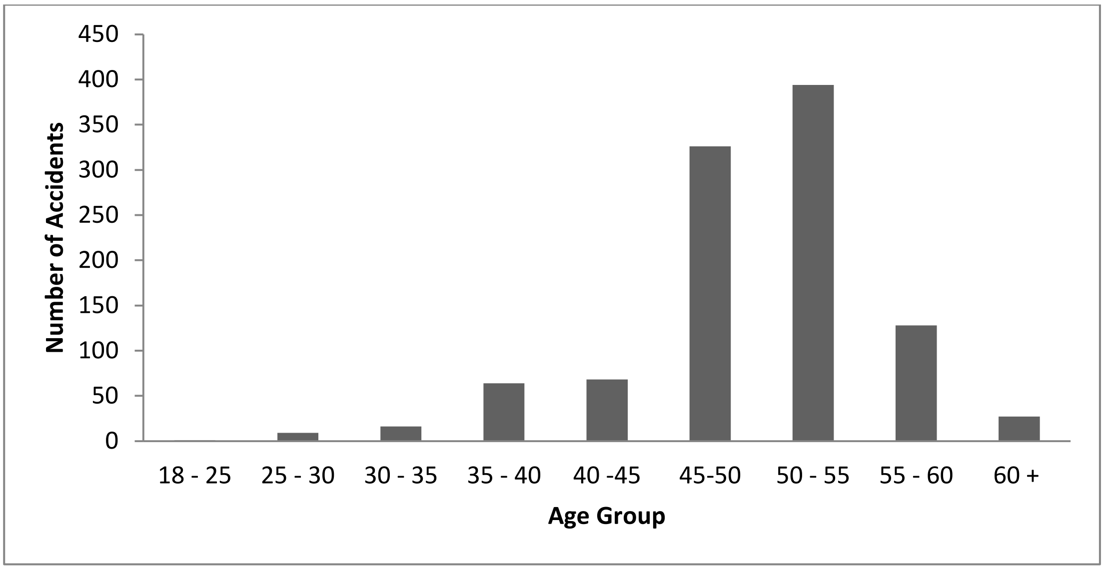


Figure 11. Distribution of accidents according to age group for ELI mine

Accident dsitribution according to age group is illustrated in Figure 11. Age group of 45-55 is dominantly higher than other ages but especially the accidents are much more higher in 50-55 of age group.

It can be seen from Figure 12 that most of the accidents in ELI mines happened because of equipment related dominantly machinery (43%) and other equipment related accidents with manual handling works follows it.

Most of the accidents happened at first shift (08-16) may probably due to the fact that all maintenance related works carried out in this shift in addition to the usual production work. Hence the number of workers are more and the probability of the accident is more.

The most affected organs due to accidents is the hands, 28%, foot with a 25 % and follows them body with 21%.

According to Figure 15, it can be seen that most of the accidents occur at spring time summer and autumn follows it. Winter has very low effect on accidents which is unexpected due to heavy weather conditions could effect on working conditions, but it can be explained that maybe the production rate is lower at winter time which is directly related with accident rate.

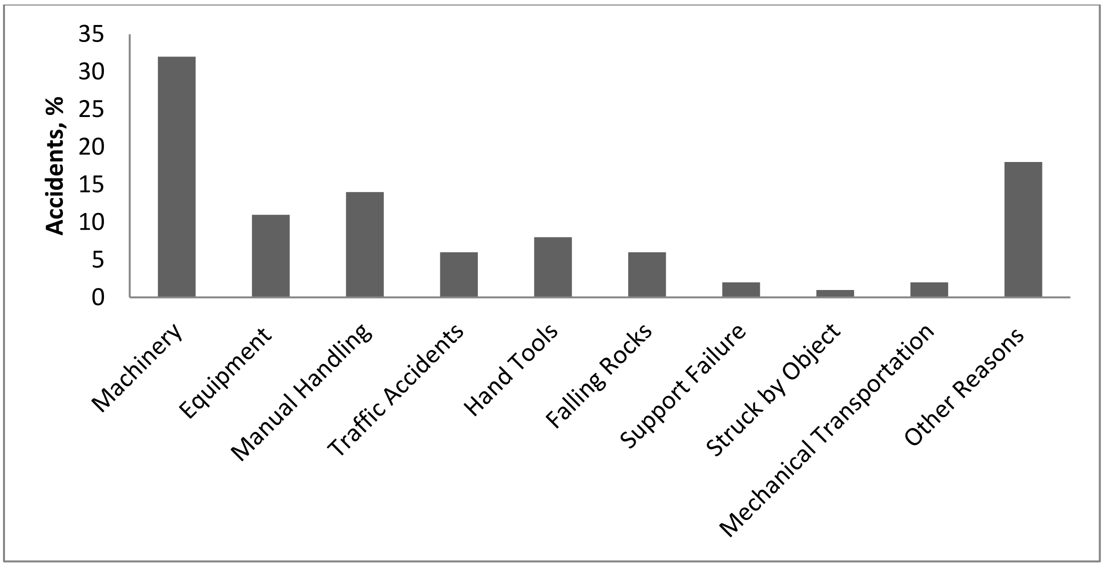


Figure 12. Distribution of reason of accident for ELI mine

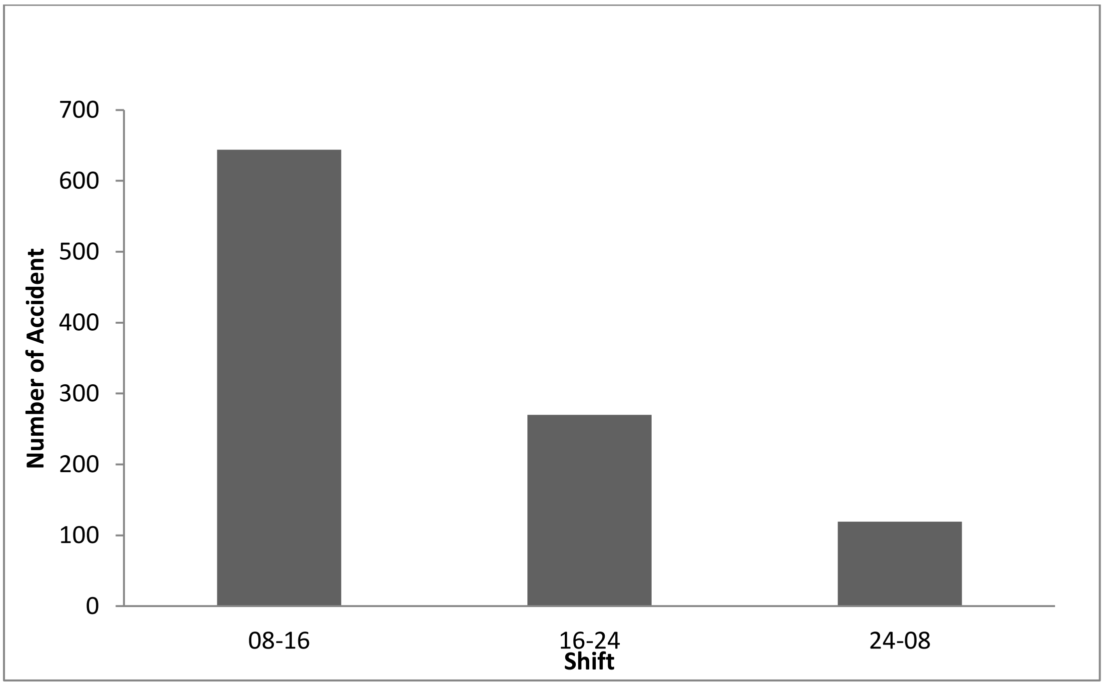


Figure 13. Distribution of accident according to shifts for ELI mine

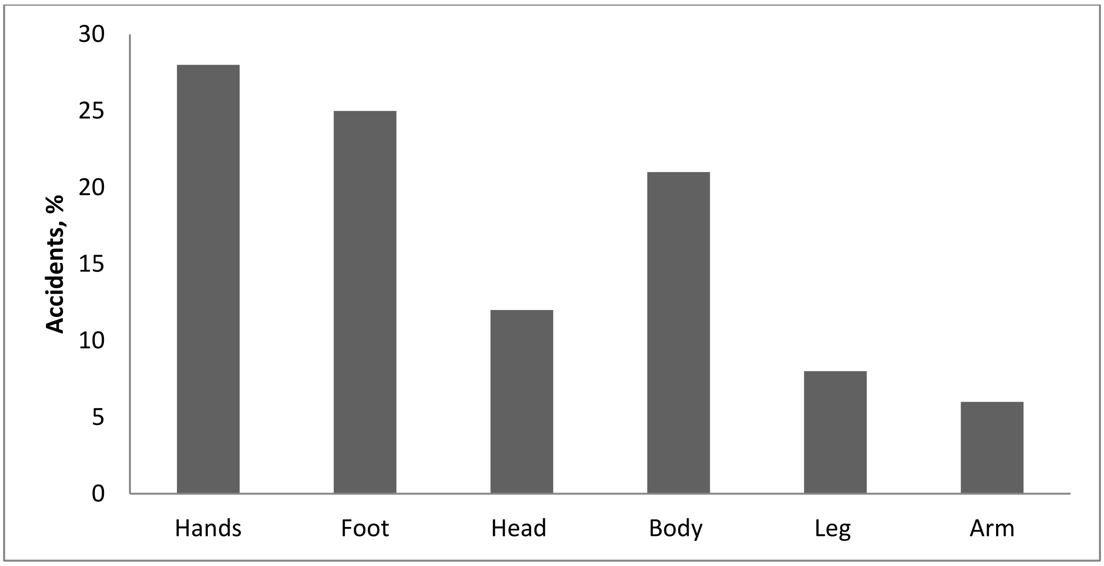


Figure 14. Distribution of accidents according to affected organ for ELI mine

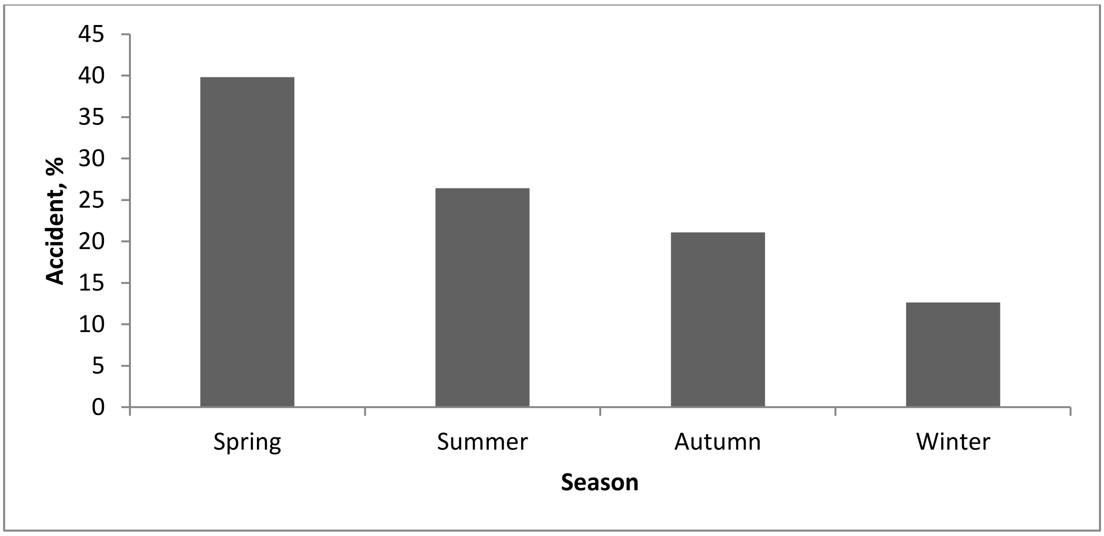


Figure 15. Distribution of accidents according to season for ELI mine

5.2. GLI Tuncbilek Underground Coal Mine Statistics

GLI Tuncbilek coal mine is one of the lignite reserves which is located in the west of Turkey in the region of Kütahya. There are some open pit lignite mines beside underground mines in the area. Conventional and fully mechanized longwall mining methods are used in the mine.

Total 1053 accident data of January 1997 – December 2011 was used in the study. The data covers, again workday loses, affected organs, reason of the accident, shift, date of accident (season) and age of the victim.

The analyses results presented in Figures 16 to 20.
According to descriptive statistics; especially workers age of 45-55 has a high trend in the accidents.Figure 16 shows the number of accidents according to age group.

It can be seen from Figure 17 that most of the accidents in GLI mines happened because of manual handling, falling rocks and equipment related items respectively.
Figure 18 shows that most of the accidents in GLI mines happened at 08-16 shift, this figure is similar to GLI mines.

The most affected organs due to accidents is the hands, 26%, foot with a 24 % and follows them body with 21% which of those are very similar to ELI mine statistics.

Figure 19 explains that the most affected organ due to accidents is the hands with 25 % and foot with a 25 %.

Finally the season effect of accidents on GLI mine is similar to ELI mine. Spring has the highest effect which Summer and Autumn follows respectively.

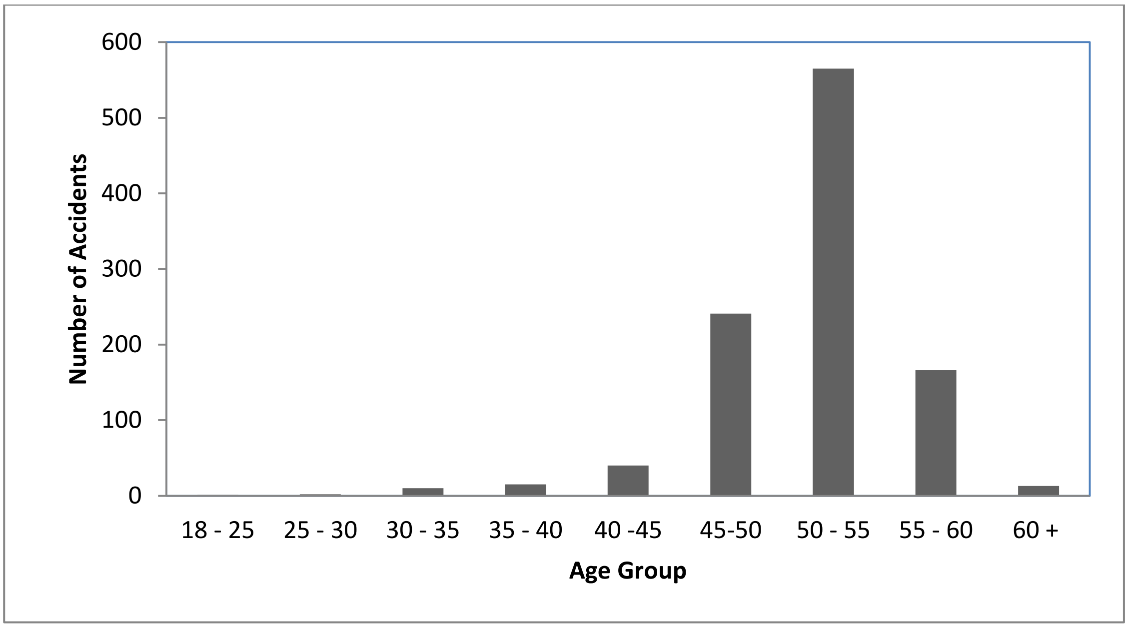


Figure 16. Distribution of accidents according to age group for GLI mine

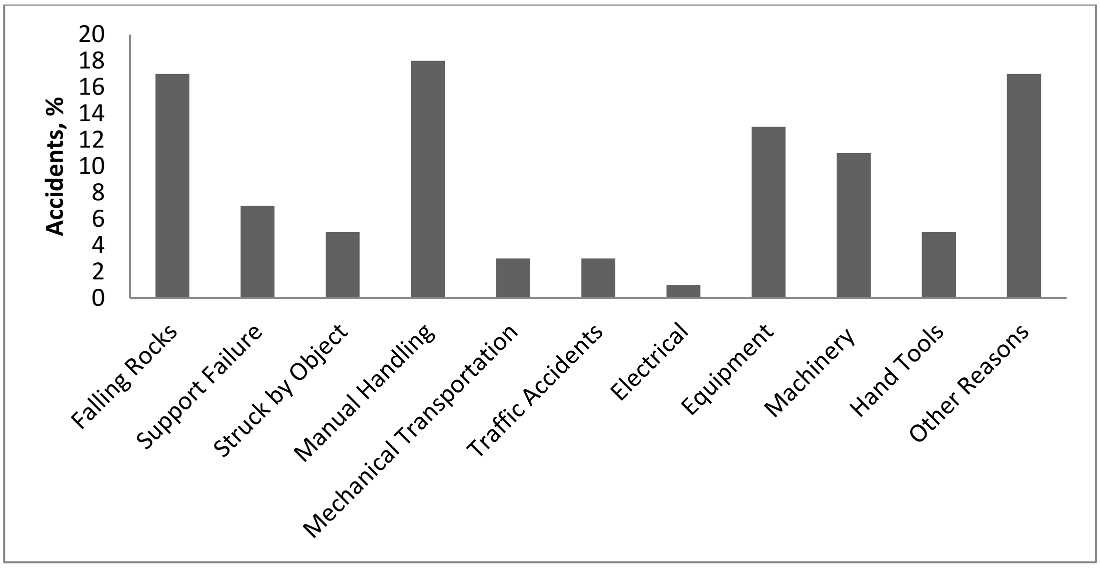


Figure 17. Distribution of reason of accident for GLI mine

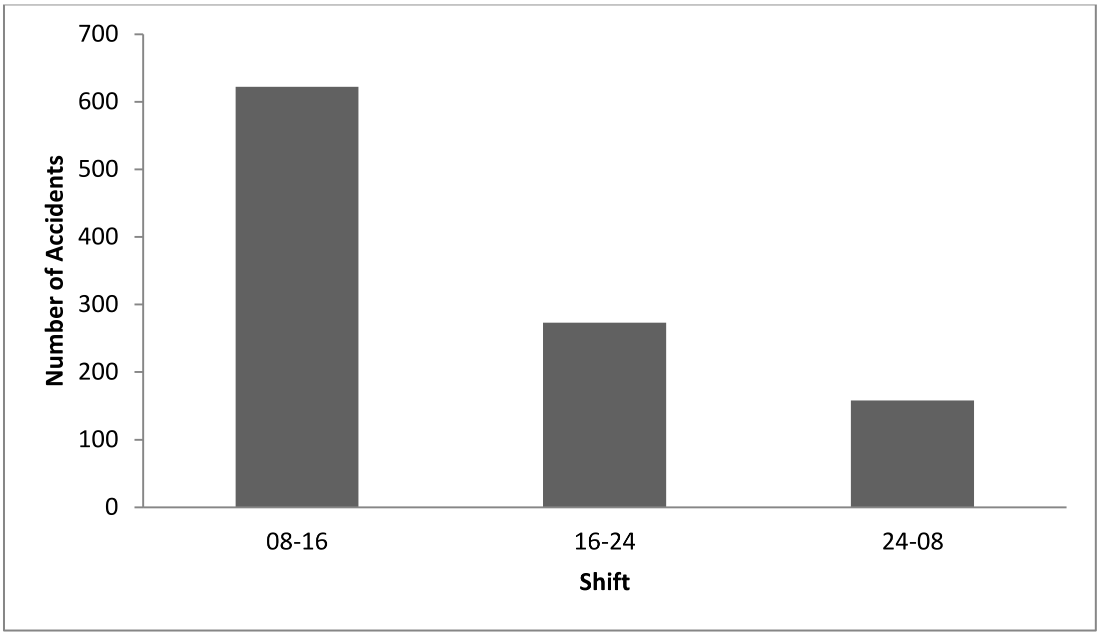


Figure 18. Distribution of accident according to shifts for GLI mine

.

Figure 19. Distribution of accidents according to affected organ for GLI mine

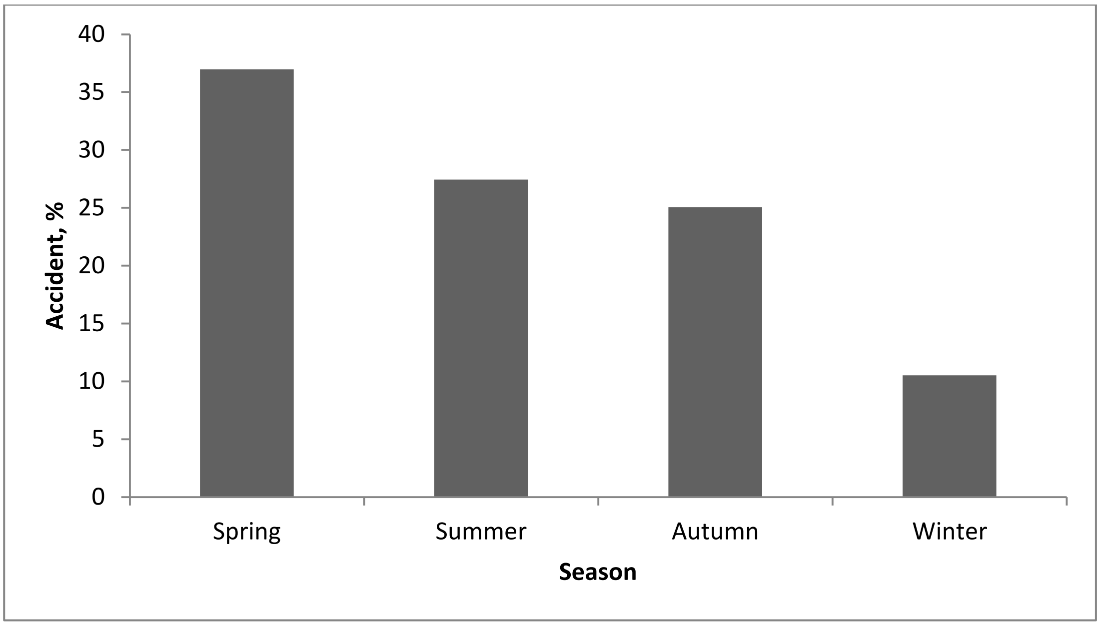


Figure 20. Distribution of accidents according to season for GLI mine

5.3. TTK Zonguldak Underground Coal Mines Statistics

TTK – Zonguldak is the only area producing hard coal in Turkey. The geological conditions are difficult and mechanized production is not possible in the mines. Generally conventional shortwall/longwall mining applied in the area. Due to the high content of methane in the coal makes the production more hazardous.

Total 8532 accident data of January 2008 – December 2011 was used in the study. The data covers, again workday loses, affected organs, reason of the accident, shift, date of accident (season) and age of the victim.

According to Figure 21; workers age of 30-35 has a high trend in the accidents which is quite different with GLI and ELI mines, and dominantly younger workers are working in TTK mines.

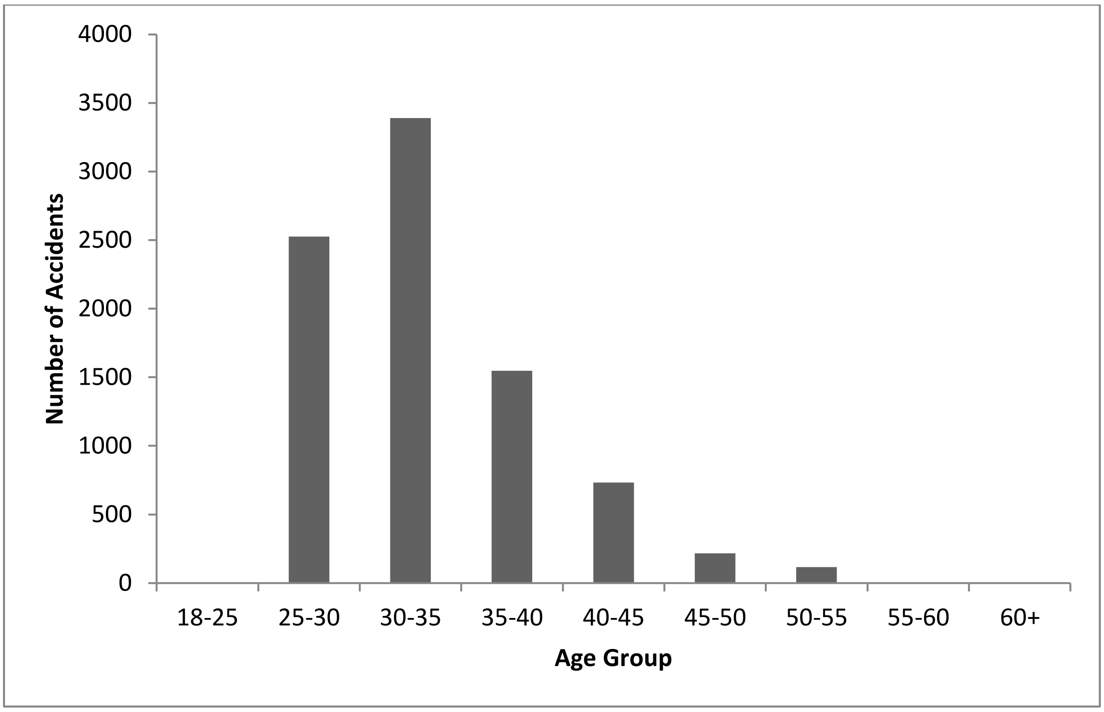


Figure 21. Distribution of accidents according to age group for TTK mines

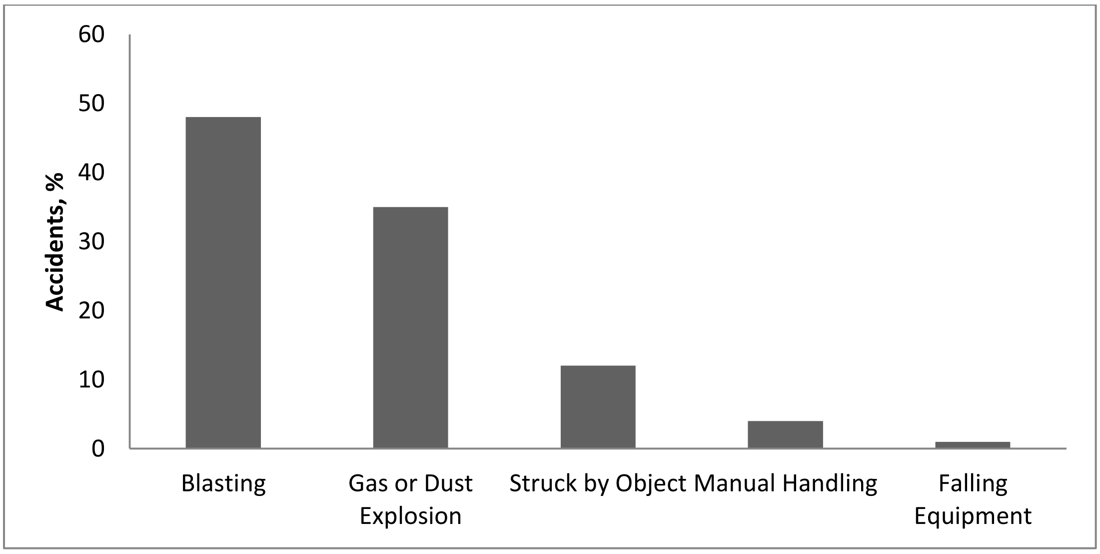


Figure 22. Distribution of reason of accident for TTK mines

It can be seen from Figure 22 that most of the accidents in TTK mines happened because of blasting and gas-dust explosion. This figure also shows that there maybe no record in the organization about minor accidents.

Most of the accidents in TTK mines also happened at 08-16 shift. This is also similar with other mine statistics as shown in Figure 23.

The the most affected organ due to accidents is arms and head with 30 % and 28 % respectively as shown in Figure 24.

Effect of season on accidents is similar to ELI and GLI mines but the only difference is summer with spring season has a high effect on accidents.

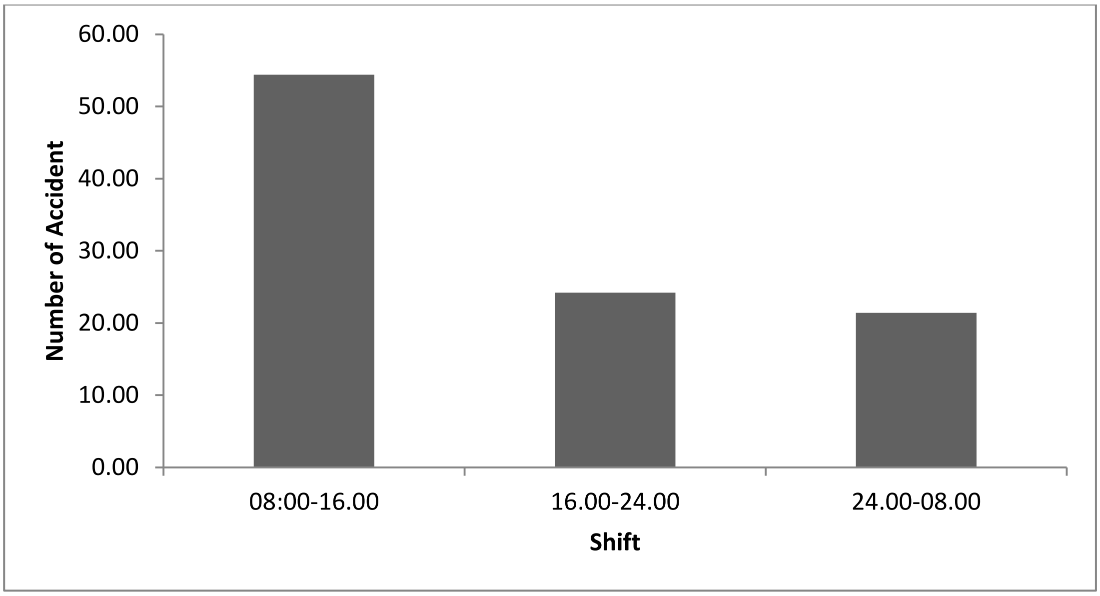


Figure 23. Distribution of accident according to shifts for TTK mines

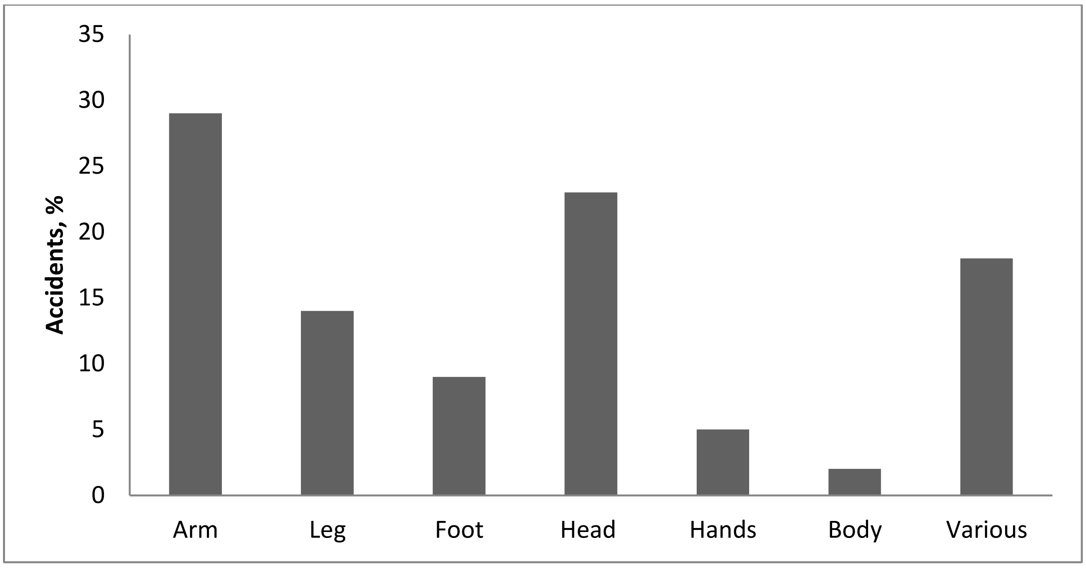


Figure 24. Distribution of accidents according to affected organ for TTK mines

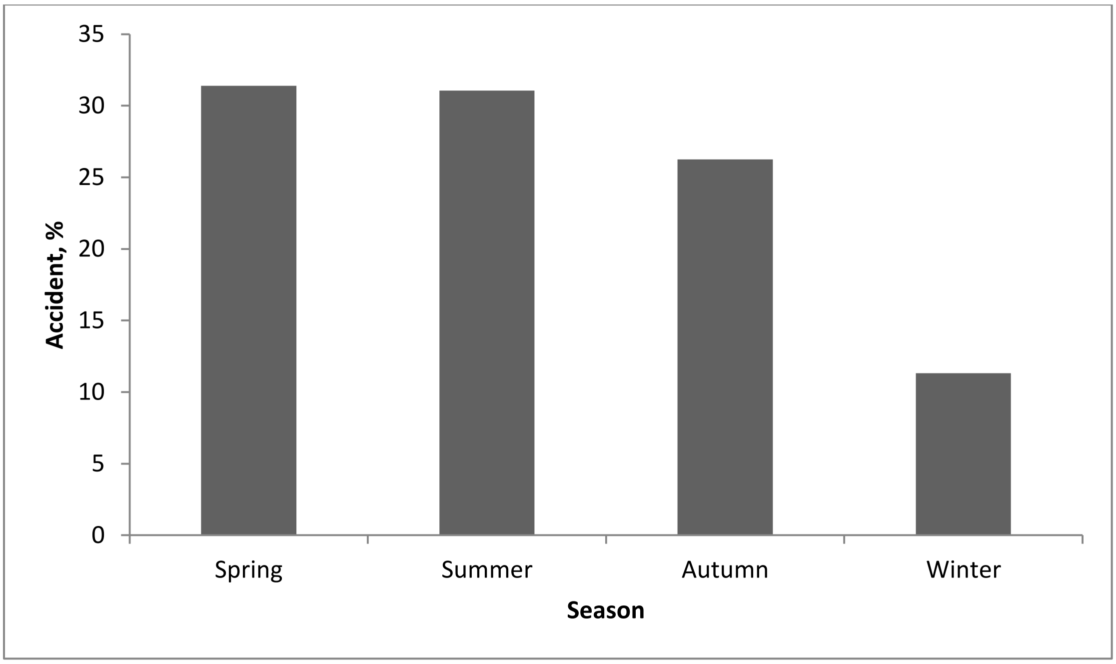


Figure 25. Distribution of accidents according to season for TTK mine

5.4. Comparison of Basic Statistics of TKI and TTK mines

When basic statistics are compared between TKI mines which are producing lignite and TTK mines which are producing hard coal, the figures are quite different which may related to complexity of the geology and the production method.

Table 16. Comparison of three mines according to age group of injuries

	Accidents, %		
Age Group	ELI	GLI	TTK
18-25	0.10	0.09	0.01
25-30	0.87	0.19	29.60
30-35	1.55	0.95	39.73
35-40	6.20	1.42	18.13
40-45	6.58	3.80	8.58
45-50	31.56	22.89	2.53
50-55	38.14	53.66	1.37
55-60	12.39	15.76	0.05
60+	2.61	1.23	0.00

Considering the effect of the age groups on accidents; in TKK mines, workers between 25 and 35 were observed as they involved accidents much more. In particular, the accident rate between 30 and 35 was 40% of the total accidents. However, in ELI mines there are more accidents between 45 and 55 ages and in GLI more accidents occur between 50 and 55 ages. In ELI and GLI mines, older workers have more accident rate comparing to TTK. In this case, it can be said that younger workers are employed in TKK mines.

In terms of all causes of accidents, gas explosion and blasting related accidents are much higher than TKI mines which are the basic result of geological conditions, nature of the hard coal and applied mining method. In Zonguldak region, the area is highly tectonised, hard coal is steeply dipping till 60° and having high amount of methane, for the coal production the fragmentation method is drilling and blasting and the more blasting is carried out than lignite mines and probability of accident occurrence is high.

Table 17. Comparison of three mines according to reason of accident

	Accidents, %		
Reason	ELI	GLI	TTK
Gas Suffocation Or Poisoning	0.00	0.00	0.26
Gas Or Dust Explosion	0.00	0.00	35.19
Falling Rocks	6.49	16.71	0.23
Support Failure	2.32	6.84	1.00
Struck by Object	1.16	5.60	11.69
Blasting	0.00	0.00	47.70
Manual Handling	13.55	18.23	3.54
Mechanical Transportation	2.13	3.51	0.02
Traffic Accidents	5.71	2.66	0.38
Electrical	0.39	0.95	0.00
Equipment	10.55	12.82	0.00
Machinery	32.24	11.11	0.00
Hand Tools	7.55	4.75	0.00
All Other Injuries	17.91	16.81	0.00

Table 18. Comparison of three mines according to shift of the accident

	Accidents, %		
Shift	ELI	GLI	TTK
08:00-16.00	62.34	59.07	54.41
16.00-24.00	26.14	25.93	24.18
24.00-08.00	11.52	15.00	21.40

According to Table 18; most of the accident occurrence time is the same and mostly accidents happen in shift 08.00-16.00 in all mines.

Table 19. Comparison of three mines according to affected organ

	Accidents, %		
Affected Organ	ELI	GLI	TTK
Head	11.71	14.53	23.15
Hands	27.49	26.21	4.47
Foot	24.88	23.55	8.52
Arm	6.10	4.46	29.42
Leg	7.94	4.46	14.28
Body	21.49	21.37	2.31
Various	0.39	5.41	17.85

When it is compared to affected organ at accidents in TTK mines the dominant organs are arms and head injuries while foots and hands are dominant in lignite mines. High and steep faces in TTK mines can cause first head and arms.

Table 20. Comparison of three mines according to season of the accident

	Accidents, %		
Season	ELI	GLI	TTK
Spring	39.82	36.97	31.38
Summer	26.43	27.45	31.05
Autumn	21.10	25.07	26.24
Winter	12.65	10.51	11.32

In all mines spring season has a higher effect on accidents. In summer and autumn time there are also accidents, but there are no comments can be made from the results. Detail research should be made in future.

CHAPTER 6

RISK ANALYSIS AND RISK MODELLING

In this chapter a risk analysis is performed by a quantitative method and past accident data is used as input. From the previous studies and researches, it is seen that a quantitative risk analysis methodology can be applied to the set of data. In the study a wider range of 14 years data is used including 4 year TTK hard coal mine data. The high number of data is very important for risk analysis studies due to understanding the effect of accidents. It is also very important to analyse the past accidents with more sampling to show the way of how studies in the workplace will trend.

From the guidance of the study of Sarı (2002) and Fine and Kinney (1976) Risk assessment methodology which Marhavilas (2009) used in the study about risk estimation in construction works adopted for ELI, GLI and TTK mines.

According to Marhavilas (2009); Risk can be quantified and be measured by a mathematical relation which is;

$$R = P.S.F \tag{6.1}$$

where:

R: the Risk

P: the Probability Index

S: the Severity Harm Index

F: the Frequency Index

Fine and Kinney (1976), proposed the method with determining the possible consequences as; catasphoric, disaster, very serious, serious, important and noticeable while Marhavillas (2009) used gradation of severity of harm index using deaths, permanent incapabilities, absence from work, etc. The current data used in this research has no evidence about the the consequences of the accidents like Marhavillas (2009) defined so the severity gradation for the research changed using absence from work as defined in Table 21.

Table 21. Gradation of the Severity of Harm Index in association with the undesirable event with the current data

Severity of Harm Index (S)	Description of Undesirable Event
10	More than 120 days absence from the work
9	Absence from work 90 days to 119 days
8	Absence from work 60 days to 89 days
7	Absence from work 30 days to 59 days
6	Absence from work 21 days to 29 days
5	Absence from work 15 days to 20 days
4	Absence from work 10 days to 14 days
3	Absence from work 5 days to 9 days
2	Absence from work 2 days to 5 days
1	Absence from work 1 day or no absence.

The Frequency Index (F) shows the number of accidents during a definite time period. In order to calculate the accidents' frequency (per day), data for 10 year time period (assumed 50 working weeks and each working week with7 working days) in the relation:

F= Number of accidents / (50x7x10) (6.2)

The equation 6.3 is used for ELI and GLI mines for availability of 10 year data and it is changed as follows for TTK mines due to availability of 4 year data.

F = Number of accidents / (50x7x4)

Using the above equation 6.1. The Risk values of ELI, GLI and TTK mines are calculated in Table 22 to 33

The Probability Index which Marhavillas (2009) proposed for calculation of accidents for specific for the group by using the corresponding number of accidents and the equation is;

P = (Number of accidents for specific group / Total number of accidents) x 10 (6.3)

The Risk Estimation Results considering reason of the accident, shift when the accident happened, age of the injured person and affected organ for ELI, GLI and TTK mines are presented below. The severity index is estimated for the worst case of the specific accident using loss workday cases and comparing with the Table 21.

Table 22. Risk Estimation of ELI mine considering Reason of the accident

	Number of Accidents	Probability Index (P)	Severity Index (S)	Frequency Index (F)	Risk Value (R)
Gas Suffocation Or Poisoning	0	0	0	0	0
Gas Or Dust Explosion	0	0	0	0	0
Falling Rocks	67	0.648596321	9	0.01914286	0.111743881
Support Failure	24	0.232333011	7	0.00685714	0.011151985
Struck By/Against Object	12	0.116166505	4	0.00342857	0.001593141
Blasting	0	0	0	0	0
Manual Handling	140	1.355275895	9	0.04	0.487899322
Mechanical Transportation	22	0.212971926	8	0.00628571	0.010709445
Traffic Accidents	59	0.571151985	10	0.01685714	0.096279906
Electrical	4	0.038722168	4	0.00114286	0.000177016
Equipment	109	1.05517909	8	0.03114286	0.262890333
Machinery	333	3.223620523	10	0.09514286	3.067044669
Hand Tools	78	0.755082285	8	0.02228571	0.134620384
All Other Injuries	185	1.79090029	10	0.05285714	0.946618725
Total Injury	1033	10	-	-	-

Table 23. Risk Estimation of ELI mine considering working shifts which accident happened.

Shift	Number of Accidents	Probability Index (P)	Severity Index (S)	Frequency Index (F)	Risk Value (R)
1	644	6.234269119	10	0.184	11.47105518
2	270	2.61374637	10	0.07714285	2.016318628
3	119	1.151984511	10	0.034	0.391674734
Total	1033	-	-	-	-

Table 24. Risk Estimation of ELI mine considering age of the injured.

Age	Number of Accidents	Probability Index (P)	Severity Index (S)	Frequency Index (F)	Risk Value (R)
18-25	1	0.009680542	1	0.000285714	2.76587E-06
25-35	25	0.242013553	9	0.007142857	0.015558014
35-45	132	1.277831559	10	0.037714286	0.481925045
45-55	720	6.969990319	10	0.205714286	14.3382658
55+	155	1.500484027	10	0.044285714	0.664500069
Total	1033	-	-	-	-

Table 25. Risk Estimation of ELI mine considering affected organ.

Affected Organ	Number of Accidents	Probability Index (P)	Severity Index (S)	Frequency Index (F)	Risk Value (R)
Head	121	1.171345595	8	0.034571429	0.323960725
Hands	284	2.749273959	9	0.081142857	2.007755497
Foot	257	2.487899322	10	0.073428571	1.826828931
Arm	63	0.609874153	10	0.018	0.109777348
Leg	82	0.793804453	10	0.023428571	0.185977043
Body	222	2.149080348	10	0.063428571	1.363130964
Various	4	0.038722168	8	0.001142857	0.000354031

As seen in Table 22 machinery, manual handling and equipment related other injuries and falling rocks are highest risk order comparing other reasons. These numbers are quite similar to the study of Sarı (2002). The only difference can be seen with machinery which can be classified as new risk in the mine.

The first shift (8-16) has the highest risk which is also obvious from simple statistics. Older employees (45-55) are trend to have a much more risk than the others. This can also be explained that the workplace has high number of older workers.

The highest risks of affected organ are hands, foot and body which are completely same with the risk assessment of Sarı (2002).

Table 26. Risk Estimation of GLI mine considering Reason of the accident

	Number of Accidents	Probability Index (P)	Severity Index (S)	Frequency Index (F)	Risk Value (R)
Gas Suffocation Or Poisoning	0	0	0	0	0
Gas Or Dust Explosion	0	0	0	0	0
Falling Rocks	176	1.671415005	10	0.050285714	0.840482974
Support Failure	72	0.683760684	10	0.020571429	0.140659341
Struck By/Against Object	59	0.560303894	10	0.016857143	0.094451228
Blasting	0	0	0	0	0
Manual Handling	192	1.823361823	9	0.054857143	0.90021978
Mechanical Transportation	37	0.351377018	10	0.010571429	0.03714557
Traffic Accidents	28	0.265906933	9	0.008	0.019145299
Electrical	10	0.094966762	4	0.002857143	0.001085334
Equipment	135	1.282051282	10	0.038571429	0.494505495
Machinery	117	1.111111111	10	0.033428571	0.371428571
Hand Tools	50	0.474833808	7	0.014285714	0.047483381
All Other Injuries	177	1.680911681	10	0.050571429	0.85006105

Table 27. Risk Estimation of GLI mine considering working shifts which accident happened.

Shift	Number of Accidents	Probability Index (P)	Severity Index (S)	Frequency Index (F)	Risk Value (R)
1	622	5.906932574	10	0.177714286	10.49746303
2	273	2.592592593	10	0.078	2.022222222
3	158	1.500474834	10	0.045142857	0.677357211

Table 28. Risk Estimation of GLI mine considering age of the injured.

Age	Number of Accidents	Probability Index (P)	Severity Index (S)	Frequency Index (F)	Risk Value (R)
18-25	1	0.009496676	1	0.000285714	2.71334E-06
25-35	12	0.113960114	4	0.003428571	0.001562882
35-45	55	0.522317189	8	0.015714286	0.065662732
45-55	806	7.654320988	10	0.230285714	17.62680776
55+	179	1.699905033	10	0.051142857	0.869380003

Table 29. Risk Estimation of GLI mine considering affected organ.

Affected Organ	Number of Accidents	Probability Index (P)	Severity Index (S)	Frequency Index (F)	Risk Value (R)
Head	153	1.452991453	10	0.043714286	0.635164835
Hands	276	2.621082621	10	0.078857143	2.066910867
Foot	248	2.355175689	10	0.070857143	1.668810202
Arm	47	0.44634378	9	0.013428571	0.053943834
Leg	47	0.44634378	9	0.013428571	0.053943834
Body	225	2.136752137	10	0.064285714	1.373626374
Various	56	0.531813865	10	0.016	0.085090218

According to Table 26 manual handling and falling rocks are highest risk comparing other reasons in GLI.

The first shift (8-16) has the highest risk which is also expected due to high number of workers in the shift. Older employees (45-55) trend to have a much more risk than the others. This can also be explained that the workplace has high number of older workers.

The highest risks of affected organ are hands, foot and body which are completely same with the risk assessment of Sarı (2002).

Table 30. Risk Estimation of TTK mine considering Reason of the accident

Reason	Number of Accidents	Probability Index (P)	Severity Index (S)	Frequency Index (F)	Risk Value (R)
Gas Suffocation or Poisoning	22	0.025788301	4	0.006285714	0.000648392
Gas Or Dust Explosion	3002	3.518930958	10	0.857714286	30.18237353
Falling Rocks	20	0.02344391	9	0.005714286	0.001205687
Support Failure	85	0.099636619	9	0.024285714	0.021777718
Struck By/Against Object	997	1.168678936	10	0.284857143	3.329065425
Blasting	4069	4.76966358	10	1.162571429	55.45074602
Manual Handling	302	0.354003048	9	0.086285714	0.274908652
Mechanical Transportation	2	0.002344391	0	0.000571429	0
Traffic Accidents	32	0.037510257	7	0.009142857	0.002400656
Electrical	0	0	0	0	0
Equipment	0	0	0	0	0
Machinery	0	0	0	0	0
Hand Tools	0	0	0	0	0
All Other Injuries	0	0	0	0	0

Table 31. Risk Estimation of TTK mine considering working shifts which accident happened.

Shift	Number of Accidents	Probability Index (P)	Severity Index (S)	Frequency Index (F)	Risk Value (R)
1	4642	5.441331614	10	3.315714286	180.4190097
2	2063	2.418239362	10	1.473571429	35.63448432
3	1826	2.140429024	10	1.304285714	27.91730998

Table 32. Risk Estimation of TTK mine considering age of the injured.

Age	Number of Accidents	Probability Index (P)	Severity Index (S)	Frequency Index (F)	Risk Value (R)
18-25	1	0.001172196	0	0.000285714	0
25-35	5914	6.932364318	10	1.689714286	117.1371502
35-45	2279	2.671433595	10	0.651142857	17.39484904
45-55	333	0.390341109	9	0.095142857	0.334243515
55+	4	0.004688782	10	0.001142857	5.35861E-05

Table 33. Risk Estimation of TTK mine considering affected organ.

Affected Organ	Number of Accidents	Probability Index (P)	Severity Index (S)	Frequency Index (F)	Risk Value (R)
Head	558	2.249093108	10	0.159428571	3.585697012
Hands	108	0.435308343	10	0.030857143	0.134323717
Foot	196	0.790004031	10	0.056	0.442402257
Arm	748	3.014913341	10	0.213714286	6.443300512
Leg	374	1.507456671	10	0.106857143	1.610825128
Body	57	0.22974607	7	0.016285714	0.026191052
Various	374	1.507456671	10	0.106857143	1.610825128
Death	66	0.266021765	10	0.018857143	0.050164104

According to Table 30 TTK risk reasons are quite different with GLI and ELI. Gas/dust explosions and blasting materials accidents have a huge risk in the mine which is expected because of the nature of the mines. The first shift (8-16) has the highest risk which is also similar with the other mines. In TTK, younger employees (25-35) are trend to have a much more risk than the others which can be defined that most of the workers are between these ages. Arms and head are the most affected organs in TTK.

The results of risk assessment are quite similar with ELI and GLI which is expected because of the production methodology and similarities of the area with same management.

Table 34. Comparison of risk assessment between mines

Variable	ELI	GLI	TTK
Reason	Machinery	Manual Handling	Blasting
Age	45-55	45-55	25-35
Shift	1	1	1
Affected Organ	Hands	Hands	Arms

TTK risk analysis results are quite different comparing with lignite mines. When items are analysed separately, blasting and gas/dust explosions have a huge impact on risks. These reasons are not present in ELI and GLI mines. This risk can be a result of a single accident which has a high severity impact with many deaths and injuries. Considering the geological conditions and type of coal (hard coal with high methane content), it is an expected risk. Although, the risk level is quite low and shown as immediate action is not necessary according to Table 13 which is Gradation of the Risk Value in association with the urgency level of required actions, the consequences of such event has a high impact on injuries and equipment loses. The management of mine should take severe precautions on the risk.

CHAPTER 7

METHODOLOGY

7.1 Pre-Processing of Data

A set of coal/lignite files was prepared, which contained the tabulated data for workday loss due to accidents in a coal mine, and the corresponding values of six mutually independent parameters affecting it for each of the three mines (GLI, ELI and TTK), which is shown in Table 35. Data was qualitative analysed and processed to be made ready for the next step, quantitative analysis. Following quantities were recorded against workday loss: Date of Birth (Injured), Date of Accident, Time of Accident (Shift), Reason of Accident, Affected Organ, Loss of Workday - in six columns. Statistics were available for a total of around thousand points each for first ELİ and GLİ and 8,500 points for TTK. It could be verified that all cause variables (hereafter mentioned as x) are independent of one another. However, the data needs to satisfy certain conditions, before it is fit for further processing:

1. A variable is continuous, provided that (i) its points should be related to each other by 'interval scale',(ii) it should contain some (at least 4) distinct values, because such values act as points for curve-fitting.

2. A variable is discrete, provided that (i) all its points should be mutually independent from one another, and (ii) there should be at least 30 instances of occurrences for each of its values. All such discrete variables need to be converted to another form before methods can be implemented on them

As per the conditions, reason, affected organ and shift could be classified as discrete variables. Date of birth cannot be related by interval scale, but if converted to age (in years), it becomes a continuous variable. Similarly, it is very likely that there would be less than 30 accidents for a given date of accident, but the problem gets resolved when this date is converted to season instead. Hereafter, age, season, shift, reason, affected organ will also be referred to, by ‘x1’, ‘x2’, ‘x3’, ‘x4’ & ‘x5’ respectively, and workday loss, a continuous variable which is the result of x1-x5, will also be referred to, by ‘y’. This can be summarized, as given in Table 35.

Table 35. Pre-processing analysis of input data

Field/Variable	Type of Variable		Conversion
	Dependent / Independent	Continous/ Discrete	
Date of Birth (Age)	Independent	Continuous	Date of birth was converted to Age (number of years)
Date of accident (season)	Independent	Discrete	Date was converted to season for classification purpose
Time of Accident (Shift)	Independent	Discrete	Data already provided in classified form
Reason of Accident	Independent	Discrete	Though provided in classified form, some sub-classes were grouped to single class
Affected Organ	Independent	Discrete	Already provided in classified form
Workday Loss	Dependent	Continuous	Result of all independent factors – already provided in a processable format (number of days)

For discrete data as per the new set of variables, a final conversion was required to convert everything in a format similar to continuous data, so that regression and neural network methods could be implemented. One such method of doing that is to convert the data to dummy binary variables.

The converted data was placed on an excel-sheet as given in Table 36.

Table 36. Set of variables provided on sheet

Age (Injured)	Reason (1-14)
Season of Accident (1-4)	Affected Organ (1-8)
Time of Accident (Shift : 1-3)	Workday Loss

For regression and neural network analysis, dummy variables could be combined with continuous ones to obtain a single equation/correlation coefficient. For fuzzy logic, they have to be analysed in their numbered form (As per table 37) because they cannot be converted to dummy variables and combined with continuous ones. Also, since there is only one continuous variable, fuzzy logic analysis will be done only for discrete variables.

Table 37. Number of variables for sites (highest value in database)

Mine	Age (cont.)	Season	Shift	Reason	Affected Organ	Total
ELI	1	4	3	14	7(9)	28 (31)
GLI	1	4	3	14	8(9)	29 (31)
TTK	1	4	3	9(14)	9	25(31)

Then the data of the total number of variables produces; number of continuous parameter + Highest value of season in database +Highest value of shifts + Highest value of reason + Highest value of affected organ, which comes at 28 for ELI, and 29 for GLI, and 25 for TTK. Maximum possible number of variables = 1+4+3+14+9 =31

7.2 Probability Analysis of Three Mines

Hazard Rate, λ: According to past accidents this rate explains the number of accidents expected in a time.

When ELI mine considered;

$\lambda_{(week)}$ = 4.64 means rate of expected accidents in 1 week.

$\lambda_{(day)}$ = 0.66 means rate of expected accidents in 1 day.

Reliability: The probability of mine's accident-free operation for a desired period (t) is named as reliability of safety. This value can be derived from equation 5.1 as putting in x=0 and remaining part of the equation $e^{-\lambda t}$ is equal to the probability of zero accident.

According to this definition;

Zero accident probability for 1 week is calculated as (t=1) = $e^{-\lambda t}$ = e^{-066x7} = 0.0098

Zero accident probability for 1 day is calculated as (t=1) = $e^{-\lambda t}$ = e^{-066x1} = 0.516

Risk of the probability of at least one accident in a time is equals to 1- $e^{-\lambda t}$

Then,

The probability of at least one accident in a week is 1-0098 = 0.9902

The probability of at least one accident in a day is 1-0516 = 0.484

Mean time between accidents, β: It is defined as the expected time between accidents and equals to $1/\lambda$ = 1/0.66 = 1.515 days

Using this systematic approach, probability analysis are made for all mines and comparison is given in Table 38.

Table 38. Comparison of three mines according to hazard indices

Indices	ELI	GLI	TTK
Hazard Rate, l (accident/day)	0.66	0.32	5.85
Reliability in one day	0.516	0.726	0.0028
Reliability in one week	0.0098	0.106	0
Risk in one day	0.484	0.274	0.9972
Risk in one week	0.9902	0.894	1
Expected time between accidents (days)	1.515	3.125	0.17

According to the probability analysis; TTK mine which produces hard coal has a very high hazard rate compared to ELI and GLI mines which produces lignite. This can also be explained with production methods and capacity, complexity in geological conditions number of workers in the operations, difference in the production techniques.

As seen in Table 38 reliability in one day is 0.516 in ELI mine, 0.726 in GLI mine and 0.0028 in TTK mine which means accident probability is much higher in TTK mine and there are accidents every day.

Similarly in all mines there is a risk of accident every week but in TTK this risk is present every day.

In TTK mine every 4 hours there is an expected risk while this number is 36 hours in ELI and 75 hours in GLI mines.

This analysis only shows the probability of accidents and some part is related with the risk but to get a precise result, the important factor is the severity of the accident. On the other hand frequency is one other factor effecting risk in many calculations.

7.3 Quantitative Analysis using Regression

The first method to be used for quantitative analysis is ordinary least squares (OLS) regression. This method deals with simply fitting a line which 'best relates' the given set of variables (continuous + dummy) linearly.

The analysis involves the following assumptions:

- The dependent variables have no mutual dependence against one another, and errors are random.
- The three mines provide similar conditions for experimentation. This means the difference in coefficients due to the three mines would be random.

The regression equation was obtained through Matlab program, which was done by first reading the matrix from file, converted the required discrete variables to dummy, and then using the 'regress' function to get the equation. Constant term was obtained by appending a unit column to the final data.

Table 39. Regression equation for ELI which has a regression value of 0.031.

Y =	*Lost Workday*
0.0802*A	*Age*
- 2.637*S2 – 2.935*S3	*Season*
+ 1.103*T1+0.145*T2	*Shift*
+ 5.671*R3 + 9.628*R4 + 0.3872*R5 + 8.407 *R(7) + 12.18*R(8) + 11.64 * R9 + 3.091 * R11 + 7.664*R12 + 5.553*R13 + 8.141*R14	*Reason*
- 20.95*O1 - 14.08 * O2 - 13.45 *O3 - 15.54*O4 -16.01*O5 - 18.21*O6+30.36	*Organ (Affected)*

Table 40. Regression equation for GLI which has a regression value of 0.054

Y =	*Lost Workday*
0.1050*A	*Age*
- 3.655*S1 - 4.960*S2 - 4.064 * S3	*Season*
-1.655*T1 + 2.228 * T2	*Shift*
+ 8.156*R3 + 10.978 * R4 + 11.459*R5 + 6.227*R7 + 10.166*R8 + 8.419*R9 + 10.633*R11 + 10.408*R12 + 2.75*R13 + 8.158*R14	*Reason*
- 0.6671*O1 + 8.152*O2 + 12.248*O3 + 4.782*O4 + 5.355*O5 + 1.636*O6 + 16.481 *O7	*Organ (Affected)*

Table 41. Regression equation for TTK which has a regression value of 0.028.

Y =	*Lost Workday*
0.400*A	*Age*
- 0.545*S1-0.993*S2	*Season*
- 15.249*T1 -12.823*T2-13.473*T3	*Shift*
+ 5.907*R2 + 21.076*R3 + 5.709*R4 + 6.043*R5 + 6.488*R6 + 9.238*R8+ 4.963*R9	*Reason*
+ 14.827*O1 + 6.075*O2 + 11.172*O3 + 13.525*O4 + 7.118*O5 + 10.755*O7 + 3.357*O8 + 4.171 *O9	*Organ (Affected)*

where,

A = Age

Si = ith season

Ti = ith shift of time

Ri = ith reason

Oi = ith organ affected

Y = workday loss per accident

The meanings of corresponding codings which are used in regression analysis are presented in Table 42.

Table 42. The meanings corresponding to codes

Season of Accident	Code	Shift/Time of Accident	Code	Reason	Code	Affected Organ	Code
Spring	1	08:00-16.00	1	Gas Suffocation Or Poisoning	1	Head	1
Summer	2	16.00-24.00	2	Gas Or Dust Explosion	2	Hands	2
Autumn	3	24.00-08.00	3	Falling Rocks	3	Foot	3
Winter	4			Support failure	4	Arm	4
				Struck by Object	5	Leg	5
				Blasting	6	Body	6
				Manual Handling	7	Various	7
				Mechanical Transportation	8	Death	8
				Traffic Accidents	9		
				Electrical	10		
				Equipment	11		
				Machinery	12		
				Hand Tools	13		
				All Other Injuries	14		

Variables states with no accidents associated with them, or those having negligibly small coefficients, are both marked as zero. So, it cannot be assumed that absence from the equation is due to zero coefficients. In the above multi-linear regression equations for all mines, it should be noted that only the coefficients of continuous variable (age) represents the strength of correlation of that variable with the equation.

The value of regression (R^2) was found to be 0.031 for ELI, 0.054 for GLI and 0.028 for TTK which shows that the obtained equations are not strongly correlated against the given dataset. This is expected, owing to fact that the number of variables in the equation is large. However, if compared to number of values in dataset, it can be said that on an average, there are more than 30 values for each variable. This proves that the data accuracy can be enhanced using some other methods, as discussed in neural network and fuzzy logic sections.

In spite of its poor accuracy, a few inferences may be drawn from the equation generated by the method. The analysis of the equation is performed as follows:

- For continuous variable (Age), the effect can be determined by the sign of variable, negative for decreasing and positive for increasing effect. It was found that the coefficient for age had different signs for different mines. Equation from TTK showed a strong correlation of loss with age, but this dependence cannot be established only on the basis of only TTK mine data. By and large, it cannot be said that there exist any relation between age and workday loss according to regression analysis.

- For seasons, it can be said that the coefficient for spring in GLI is certainly higher than the other three seasons for ELI and TTK. It has zero coefficients for ELI and TTK, because it shows 0 values, and has a reasonably large number of observations corresponding to it. For site GLI also, its value is higher than that of others. This means that workday loss is higher (per accident) in spring season than other seasons.

- The results for ‘reason’ were not strong enough, but some reasons like Manual Handling (7) were found to have high values for all the three sites, and can therefore, be claimed to be enhancing workday loss per incidence. Other results for reason are not too commendable with regression analysis.

- ‘Affected Organ’ has provided some good results. It is clearly evident that workday loss gets minimized for head injuries (1) since this column has lowest values for ELI, GLI and TTK mines. Similarly, foot injury (3) has highest & arm injury (4) has comes after that for both mines, which proves that workday loss is higher for foot and arm injury, as compared to other injuries.

- For shift, one result could be obtained which is as follows. For second shift, the correlation value was found to be either highest (ELI or TTK) or second highest (GLI). This proves that second shift causes more workday loss than the other two shifts.

The results can be summarized as shown in Table 43.

Table 43. Effects of parameters on workday loss – results from regression

Sr. No.	Parameter	Effect on workday loss	ELI	GLI	TTK
1	Age	-	-	-	+
2	Shift	(2) 08:00-16:00 (+)	+	-	+
3	Season	(1)Spring (+)	+	+	+
4	Reason	(7) Manual handling (+)	+	+	+
5	Affected Organ	(1) Head (--), (3) Foot (++), (4) Arm (+)	+	+	+

* (--) high decreasing effect, (-) decreasing effect, (+) enhancing effect, (++) high enhancing effec

Another result which could be obtained from this method is the ‘intensity of replacement within variables’. For example, the coefficient values corresponding to seasons (2, 3) for ELI were found to be (-2.63,-2.93) respectively. Replacing one by another modifies the final result (workday loss per accident) by (+/-) 0.3. This is because dummy variables take the value 1 or 0 for availability or unavailability, which means coefficients simply add up for all such existing states. Such a difference for GLI was found to be 0.9; there is no significant result for TTK.

Differences in corresponding values for shifts (1, 2) were found to be 0.95, 3.8 and 0.65 for ELI, GLI and TTK respectively. So, it can be said that the effect of shift replacement is stronger than the effect of time replacement. This logic can be quantified by taking the median values of the obtained regression coefficients, and then comparing them against highest and lowest coefficient values for the same parameter. By doing this, it could be established that reason is the most significant parameter as per the current method, as it has the highest replacement effect for the three areas combined together. This was further verified by neural networks analysis as well, as will be discussed later.

When the analyses has no site restriction; another type of comparative analysis, which is possible in context of the current study, is the inter-site comparison. In other words, are all season/shift/reason/organ affecting workday loss in the same order, in all the three sites? To answer this question, one of our earlier assumptions of all sites being similar has to be removed. So, it is assumed till this analysis, that different sites might be influenced by same factors in different ways. This is explained in the following points:

1. Age - Effect of age was found to be most dominant in TTK. So, if aged individuals are employed at TTK, their productivity could go down.
2. Season – For GLI, summer was found to be having the least coefficient. If one site is chosen per season, GLI is recommended for summer.

3. Shift – For GLI & TTK, shift (0800-1600) was dangerous. However, for ELI, shift (2400-0800) had a higher value.
4. Reason – For ELI, most of the accidents were found to occur due to blasting. Also, struck by object was found to be least effective. In contrast, for GLI, it was hand-tools, which was the weakest cause of workday loss. Other than that, no concrete interpretations could be found on 'reason' factor'.
5. Affected Organ – Contrary to the ELI and GLI, the coefficient for 'head' was found to be highest in TTK. Thus, TTK is much more susceptible to head injuries, as compared to ELI and GLI, for the same amount of workday loss.

7.3.1 Regression Using Modified Method

An improvement to the previous regression method was attempted through an implementation of a better algorithm – irrelevant variable removal approach. The new algorithm accounts for the fact that the removals of those variables which do not affect output (lost workdays), enable the remaining ones to produce a better quality fit, as explained before.

Its algorithm goes as follows:

1. Calculate p values of normal distribution for all variables included in the regression equation.
2. Check whether there exist at least one variable with the p value greater than 0.2. If no, stop. If yes, go to step 1.
3. Remove the variable with the highest p value.
4. Perform regression using the remaining variables.

Table 44. The equation obtained for modified regression for ELI

Y =	*Lost Workday*
-2.572*S1 -3.006*S2	*Season*
+ 3.725*R7 + 7.247* R8 + 7.166*R9 + 2.743*R12 + 3.441*R14	*Reason*
-6.986*O1 -4.166*O6	*Organ (Affected)*

Table 45. The equation obtained for modified regression for GLI

Y =	*Lost Workday*
3.960 * T2	*Shift*
+ 9.347 * R3 + 11.876 * R4 + 11.853 * R5 + 7.433 * R7 + 10.84 * R8 + 9.764 * R9 + 11.69*R11 + 11.1 * R12 + 9.558 * R14	*Reason*
+7.193 * O2 + 10.898 * O3 + 14.992 * O7	*Organ (Affected)*

Table 46. The equation obtained for modified regression for TTK

Y =	*Lost Workday*
0.319 * A	*Age*
-2.3 * T	*Shift*
+14.72 * R	*Reason*
+10.21 * O1 + 6.53 * O3 + 8.954 * O4 + 2.546 * O5 + 6.114 * O7	*Organ (Affected)*

where coefficients (A, Ti, Ri, Oi, Si) have the same meanings, as provided in quantitative analysis using regression section. Table 47 shows the eliminated P values of ELI mine for the regression.

The results for modified regression were not found to be better than those for ordinary least squares regression in our case (correlation values of 0.025, 0.049 and 0.026 against 0.031, 0.054 and 0.028 respectively). So, instead of this modified method of regression, normal regression would be used for further research and analysis. Inter-site comparison is difficult, since most of the coefficients included in the three equations are different.

Table 47. P values of eliminated variables for ELI

Variable	**Value**	**Variable**	**Value**	**Variable**	**Value**
Age (continuous)	0.417	**Reason**		**Organ**	
		Gas Suffocation Or Poisoning	S	Head	
Season of Accident		Gas Or Dust Explosion	S	Hands	0.597826
Spring (S1)	S	Falling Rocks	0.42	Foot	0.23963
Summer (S2)	N	Support Failure	0.26	Arm	0.471369
Autumn (S3)	N	Struck by Object	0.98	Leg	0.401461
Winter (S4)	N	Blasting	S	Body	N
Shift/Time of Accident		Manual Handling	N	Various	S
08:00-16.00 (T1)	0.463	Mechanical Transportation	N	Death	A
16.00-24.00 (T2)	0.952	Traffic Accidents	N		
24.00-08.00 (T3)	S	Electrical	S		
		Equipment	0.631		
		Machinery	N		
		Hand Tools	0.51		

* The symbol ‘S’ stands for ‘eliminated by system’, ‘N’ for ‘not eliminated’, A for absent. Other quantities are corresponding P values

7.4 Quantitative Analysis using Neural Networks

In spite of the simplicity of design and implementation of previous method and its good results for several parameters, its effectiveness remains questionable for the purpose of study because of the fact that it produced extremely low correlation values (0.031, 0.054, 0.028) from all three mines, so a better and more robust method was required, which could either improve regression or implements some other algorithm to produce results better than Ordinary Least Squares (OLS) regression. One such method is Neural Networks which improves the performance of regression methods by ‘training’ the function for preparing it for analysis. It does so by iterating the function’s outputs and inputs through a hidden layer of neurons., which improves the performance of regression methods by putting the desired ‘weights’ for processing regression data. Such weights are calculated using the input and output values of training set. This training set is developed from the random values from the given data. Some values are also used for creating a validation set, for verifying that the data is getting improvements with training sets. Remaining values are used for testing purpose.

Apart from inputs and outputs, there is a dynamic hidden layer, which contains weights of variables. Number of neurons decides the number of variables which store this information.

Matlab’s Neural Network tool takes the following inputs: x(causes) ,y(effect) ,n (number of neurons), distribution of experiments in training/validation/testing. The default value for training/validation/testing could be used for analysis. However, the number of neurons needed to be modified considering the huge number of variables in the Dataset. Using several possible values of neurons, it was found that the regression values could be significantly improved for ELI – from 0.10 to 0.28, as the number of neurons was increased from 10 to 100, as shown in Table 48 below. Table

48 also proves that the output is maximized when number of neurons ~ number of variables, and does not improve further on increasing the number of neurons. Increasing the neuron number improves both the efficiency and the processing time, but the former diminishes beyond a point. This is because after a certain point, equations created due to extra layers are redundant. Table 48 shows that the quality of fit improves with the number of neurons, till 20 neurons. However, two huge ‘number of layer’ values were taken to verify that increasing number of neurons does not improve the quality of fit indefinitely. For n = 50, a correlation coefficient value of 0.25 was obtained, and for n = 100 this value was 0.23. These two values are almost similar, and not too far from the one obtained at n=20 (0.28sw). Best fit is achieved with highest regression value.

Table 48. Improvement of regression value with change in number of neurons for GLI

Number of neurons	Regression Value
10	0.10
12	0.18
15	0.18
18	0.19
20	0.28
25	0.21
30	0.26
50	0.25
100	0.23

Number of neurons is chosen as 20 according to best fitting regression value and used for analysis of all mines.

Using the code, the final regression value was computed. However, it is possible to obtain regression value separately for training, validation and testing set, as shown in the Figure 26 to 28 for ELI, GLI and TTK.

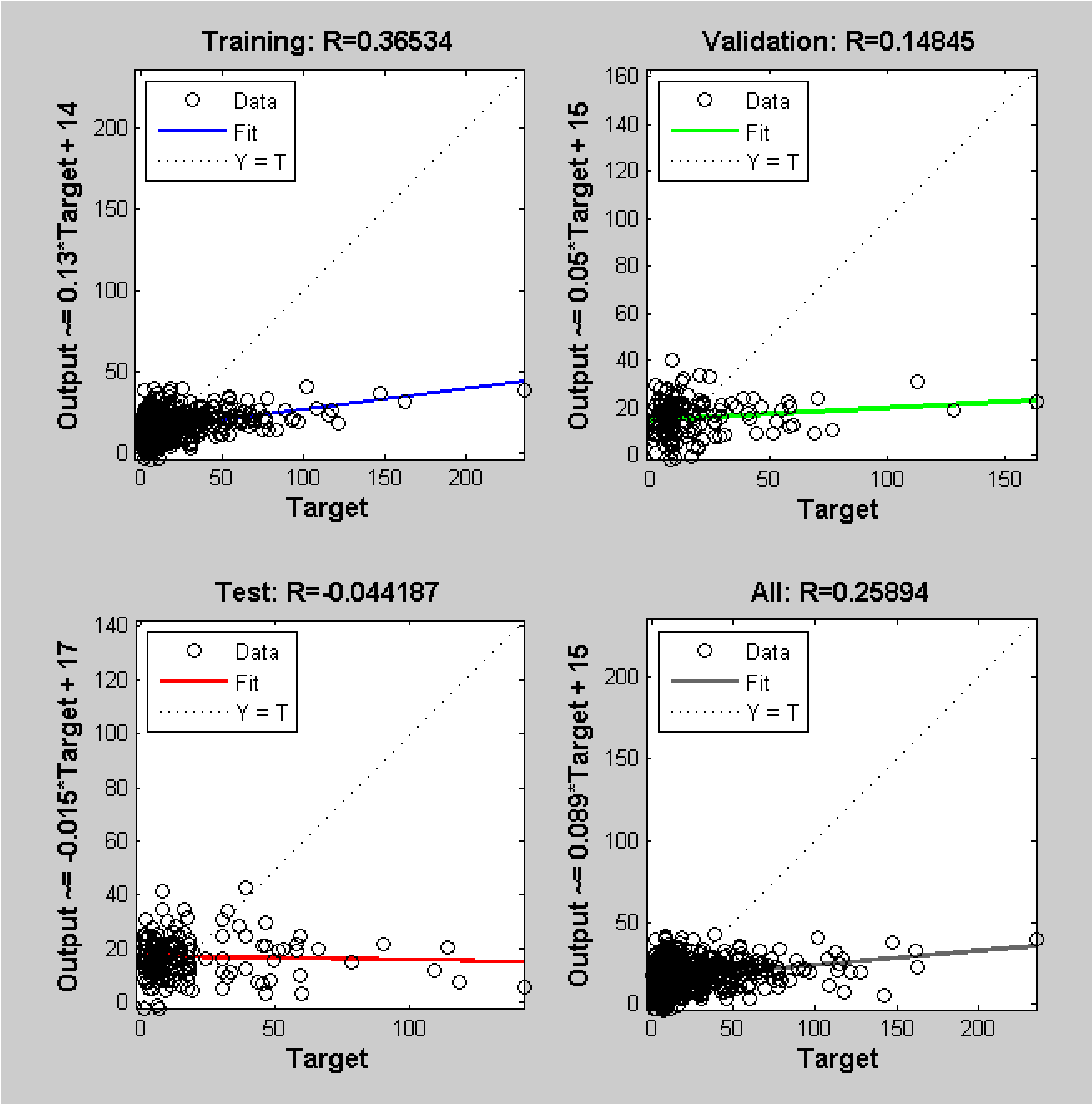


Figure 26. Regression values for ELI, 20 neurons (70% training, 15% validation, 15% testing conditions)

The method does not generate a unique set of coefficients, since it picks a random seed for every iteration.

Table 49. A typical set of values of coefficients and equation for ELI

Y = - 0.0434 * A	*Age*
- 4.716 * S2 - 5.3498 * S3	*Season*
- 2.413 * T1 - 0.4559 * T2	*Shift*
+ 4.5279 * R3 + 1.0682 * R4 + 1.5412 * R5 + 0.2593 * R7 + 3.9526 * R8 + 0.7957 * R9 + 3.8772 * R11 + 1.9106 * R12 + 1.7908 * R13 - 1.6177 * R14	*Reason*
+ 3.1942 * O1 + 1.0054 * O2 + 3.8285 * O3 + 3.7154 * O4 - 0.4717 * O5 + 3.4358 * O6	*Organ Affected*

Table 50. A typical set of values of coefficients and equation for GLI.

Y = - 0.02481 * A	*Age*
- 0.06479 * S2 – 1.34419S3	*Season*
+8.49 * T1 + 8.732 * T2 + 8.039 T3	*Shift*
+ 3.56 * R3 + 0.48 * R4 + 9.5 * R5 + 2.162 * R7 + 6.314 * R8 + 3.66 * R9 + 6.39 * R11 + 4.99 * R12 + 1.519 * R13 - 1.96 * R14	*Reason*
-4.87 * O1 + 5.908 * O2 + 8.407 * O3 + 1.912 * O4 - 0.483 * O5 – 0.439O6 + 9.348 * O7	*Organ Affected*

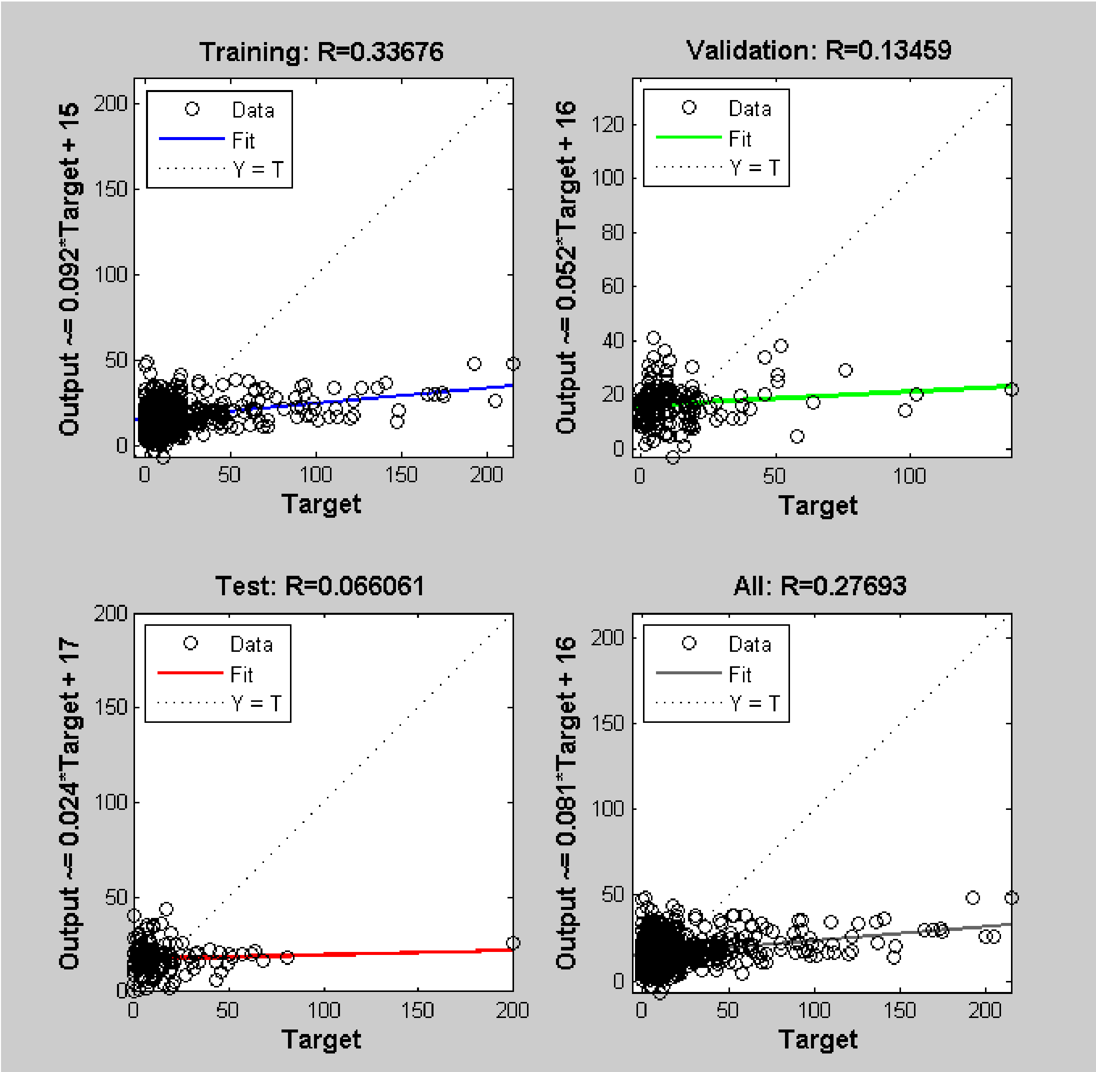


Figure 27. Regression values for GLI, 20 neurons

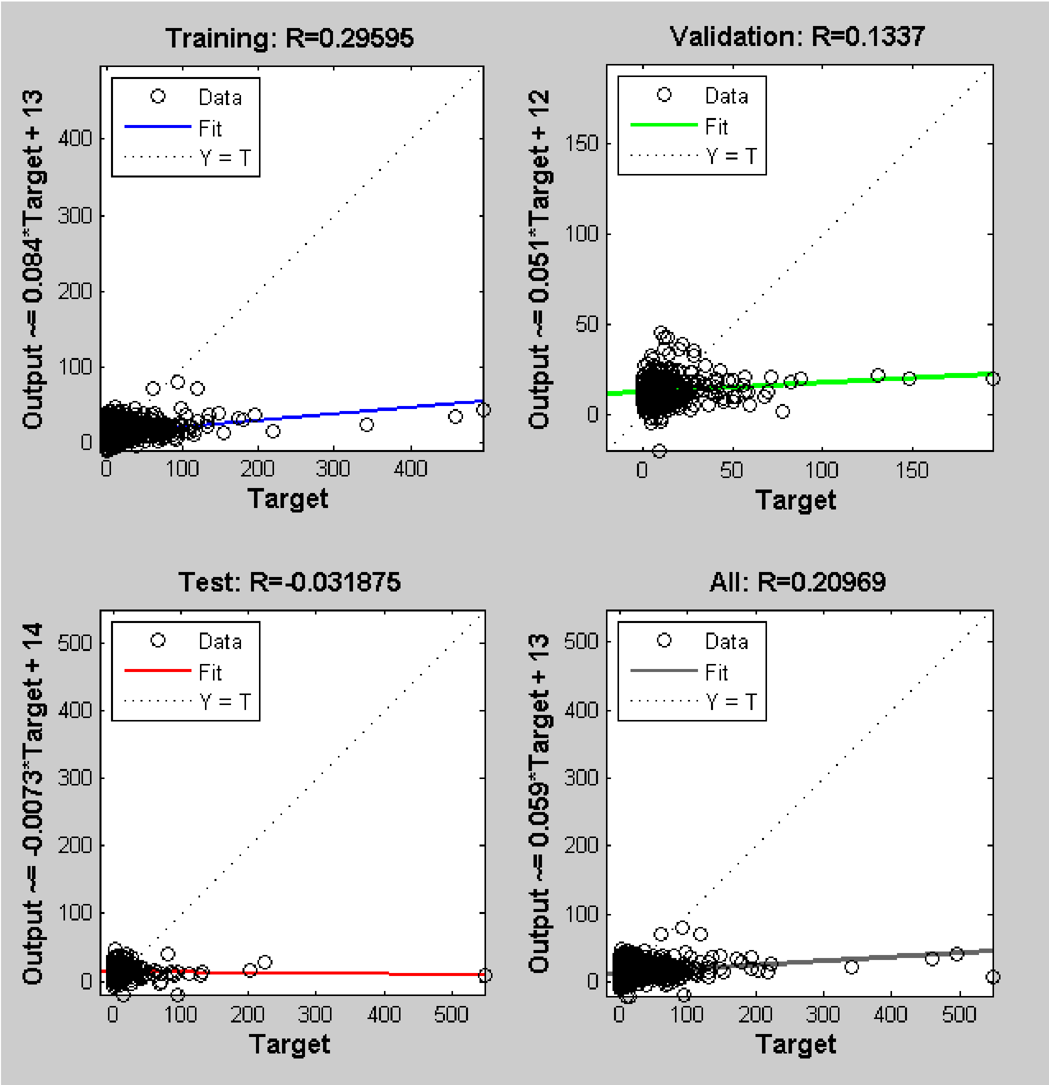


Figure 28. Regression values for TTK, 20 neurons

Table 51. A typical set of values of coefficients and equation for TTK.

Y = 0.28819 * A	*Age*
- 1.6106 * S2 + 0.0243 *S3	*Season*
-0.1869 * T1 + 4.463 * T2 + 4.086 T3	*Shift*
+1.426 * R2 + 1.227 * R3 - 5.778 * R4 + 1.411 * R5 – 0.475 * R6 - 3.974 * R7 – 0.559 * R9	*Reason*
8.069 * O1 + 5.726 * O2 + 6.056 * O3 + 8.035 * O4 + 6.016 * O5 + 6.644 *O7 + 2.895 * O8 + 4.088 * O9	*Organ Affected*

Though neural network method improves regression (R^2=0.28) at the expense of increased complexity and time required for the machine, its final results will have the same limitations as that of regression – the quality deteriorates as the number of inputs increase. This is because similar to regression; neural network method also relies on finding the best fit solution for all input variables taken simultaneously, and minimization of the corresponding error. This algorithm does not fully exploit the advantage of having a huge number of points in the database.

Results similar to regression for mines may be found in neural networks method as well. However, it is difficult to comment using such results, since the coefficients generated by the method depend on seed value, that is, they are random. The following are a few results for inter-site comparison for one such run:

1. Age - Since age factor (a continuous variable) has negligible coefficient for ELI and GLI, and a significantly high value for TTK, it can be said that young individuals should be employed to TTK – this is similar to what was found in regression.

2. Season – Contrary to regression, summer was found to have the least coefficient for TTK. Thus, this method recommends TTK for summers, for the same workday loss.
3. Shift – 0800-1600 shift has effect on workday loses in ELI and GLI but in TTK it is 0000-0800 shift which is completely different with regression analysis, in regression the effect of shift 0800-1600 was higher for GLI and TTK.
4. Reason – Manual handling was not found to be an enhancing factor for TTK but it has major effect in GLI and ELI mines. Thus, if both regression and neural networks methods are combined, it can be said that blasting is an enhancing factor for TTK.
5. Affected Organ – Similar to results obtained from regression, head injuries resulted in higher number of workday loss as compared to other injuries for TTK. Thus, this is a definite result.

This method produced some results in addition to those predicted by simple regression. For example, if the files are compared for time effect, it can be seen that accidents occurring in shift 0800-1600 have more severe impact than those in shift 2400-0800 in ELI and TTK but shift 1600-2400 has more effect for accidents in GLI comparing to others

Most of the previous results can be validated using this method. For example, the effect of manual handling can still be found to be 3rd (ELI) or 4th out of 10 values for reasons. This means the given reason definitely results in severe accidents. Similarly, for affected organ variable, head injury effect was found to be lower (4/6 – ELI) and least severe (GLI). Foot injury effect was confirmed to be highest/second highest respectively, and arm injury – second highest respectively. Age (continuous parameter) still had different signs for two sites. So the new results can be tabulated as provided in Table 52. However, in this case, one more factor could be found in addition to manual handling – blasting, whose effect was nearly highest for TTK.

Table 52. Results from neural analysis, and their comparison against regression method

Parameter	Effect on workday loss		
	ELI	GLI	TTK
Age	None	None	25-35 (+)
Shift	0800-1600 (+)	0800-1600 (+)	0000-0800 (+)
Season	Spring (+)	Spring (+)	Summer (+)
Reason	Manual Handling (+)	Manual Handling (+)	Blasting Material (+)
Affected Organ	Foot (++), Arm (+)	Foot (++), Arm (+)	Foot (++), Arm (+)

Using exactly the same method as that for OLS regression, similar results for replacement effects were established for neural networks also. But on this occasion, some different weights were used to derive the values, so their intensities do not result in direct contribution to workday loss like in regression.

7.5 Fuzzy Logic Analysis

The following is a starting premise for formation of rules. First, every result is a summation of individual efforts of all the inputs, but the problem lies in measuring such individual efforts. Also, the result of the efforts (y) for a particular accident might have a purely random value, which is nowhere related to the values of inputs. Nevertheless, the following can be said with a very good accuracy: If higher than average result was found in presence of a particular state of some discrete variable, it is more likely that if a particular observation has that value, it will have higher than average workday loss. For example, if most of the accidents corresponding to the season 'spring' have high values of workday loss, it can be said that spring season is responsible for increasing the severity of workday loss.

The above logic can be further extended to assume that the state of variable (such as spring season) is responsible to the extent of average of the resultant 'y' values of all the accidents where it is present. Fuzzy logic rules can be generated using this logic. However, it requires some assumptions, as given below:

- Sufficient number of points exists for every state, such that there is no bias due to the non-randomness of the pattern of other variables affecting the result whose average is represented by the considered state. Only states with more than 30 instances of occurrences will be considered for that variable. Such a state will be referred to as ‘significant’ henceforth. 30 are taken here because each state sees only four variables around it, but if the numbers of other variables were more than 4, or number of states per variable was large, this number would increase from 30.
- There is zero dependence between variables. In other words – the effect on the value of state of a particular variable should not be because of some other variable influencing it.
- The distribution of states within the variables should be, to a large extent, random. For example, if most of the cases of severe head injuries were removed from the sample, the final results will not be good.

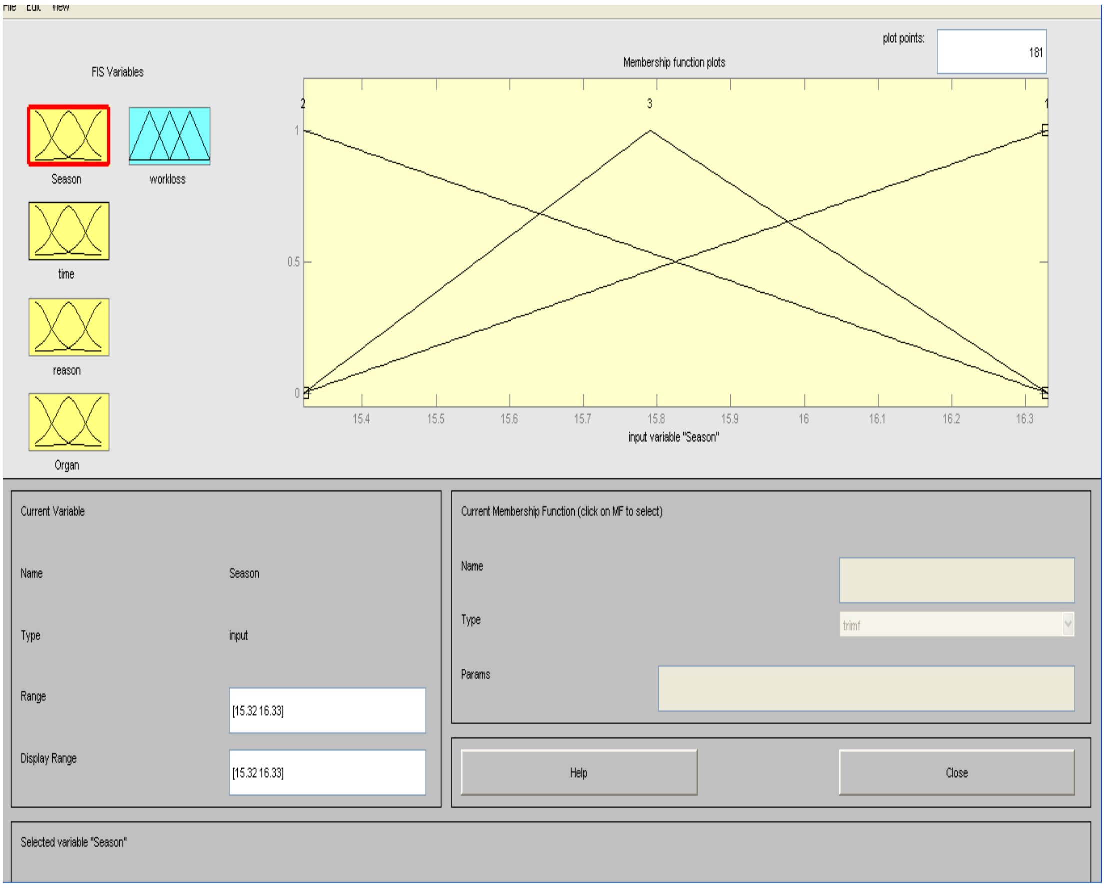


Figure 29. Membership function for 'seasons' for GLI

Having laid down the assumptions, the algorithm for creating the rules can be defined as follows:

- Select the first discrete variable and scan through its states.
- If the number of accidents including the given state is less than 30, exclude the state. If it is greater than 30, compute the corresponding average in the corresponding values for workday loss **('F').** For instance, if season = 1 corresponding to rows/accidents 4, 5, 19, 24 the corresponding average value will be average (F4, F5, F19, F24).
- Repeat this for all states of all variables to generate a value corresponding to each of the 'significant' states.
- Compute highest/lowest/median (H/L/M) values for variables.

- Each state will have an objective function of a triangle, defined as (H, V, L), where H = highest value for the variable, L = lowest value for the variable and V = Value corresponding that state. For such a fuzzy function, value of state corresponding to L will decrease from highest (at L) to 0 (at H) as we move to state with higher value. It is reasonable to expect this because (i) as the contribution of lower values decrease, the value of function increases (ii) variable's value equals to that of a state means the corresponding state's contribution is 100%. So, a state's fuzzy logic function member is a peak at state's value and two legs at maximum/minimum values. Figure 29 shows the member functions for season variable of GLI. (1) Corresponds to Highest value, (2) to Lowest and (3) to Median value.
- Rules are yet to be decided for result variable (workday loss) which is continuous and ungrouped. For doing that, it has to be grouped. The simplest way of doing that is to form 3 equal groups according to the order of expected values. This grouping is not done according to the actual values, since they could be random. So, the values in all groups combined together cover all the possible combination of discrete inputs. The rules are presented in Appendix 1.
- Similar method may be used for making member function for result also, even though it is a continuous variable. However, to adjust for the fact that unlike discrete functions, it has a range of values to represent a state instead of a single value, a trapezoid type function would be better suited instead of a triangle function. The only difference between that and triangle function is that it accepts four position instead of 3. Similar to inputs, the extreme values for member functions of states are lowest and highest value for low and high respectively. However, there are two values for middle functions - lowest (V_L), and the highest (V_H) values from that state. Consequently, the member function corresponding to 'low' becomes (L, L, V_H, H) for 'high' it becomes (L, V_L, H, H) and for medium, it becomes (L, V_L, V_H, H). So far, the membership functions have been defined.

- The final task is to formulate the rules using the membership functions of inputs and outputs. This is now straight forward, since combinations of inputs provide the corresponding output. For example, if workday loss = 2 (medium) for season = 1, time = 2, organ = 5, reason = 3, this becomes a fuzzy rule.
- Translate all combinations and corresponding values to rules, and generate a fuzzy model in Matlab, using those rules.

The implementation of rules was done using excel functions/macros and Matlab fuzzy logic toolbox, in the following steps:

1. First, prepare the inputs for fuzzy logic using a Matlab program which performs steps (1-3, 6), and provides two types of outputs (i) excel-sheet average value matrices corresponding to each variable in separate sheet (ii) another excel-sheet containing all combinations of inputs versus output state (low/medium/high).
2. Next, compute the values for member values for inputs and outputs on excel, and save them to a new file on fuzzy logic toolbox.
3. Last, feed the rules in the rules section of the toolbox. This immediately generates an interface which generates values for the combinations.

Since the results of this method are based on quality attributes, it cannot be quantified for comparison, like the other two methods. Also, this estimator produces output from the given values using a set of rules, rather than the set of equations. This makes the comparison even more difficult. However, the quality of data produced by fuzzy logic can be compared to the table by randomly picking the sample and applying the rule. For example, for observation 76 (1, 4, 3, 2, 3, 1), actual value of the observation was 15, while that predicted by fuzzy logic was 14.3. To a large extent, the similar observations apply to the entire dataset.

A snapshot of rules, is as shown in figure 30. The four columns on the left represent four input variables (Season, time, reason and affected organ) and the column on the right represents their result (which is equivalent to a sum of corresponding averages as discussed before, but with some fuzzy adjustments). Thin red lines on the inputs represent its current average value, which can be changed by moving it to left or right.

Apart from rule generation, it also enables users to view the effects of inputs taken 2 at a time, which can be viewed from 'surface'. However, no commendable observations could be found from surface analysis due to it shows only 2 input variables with output.

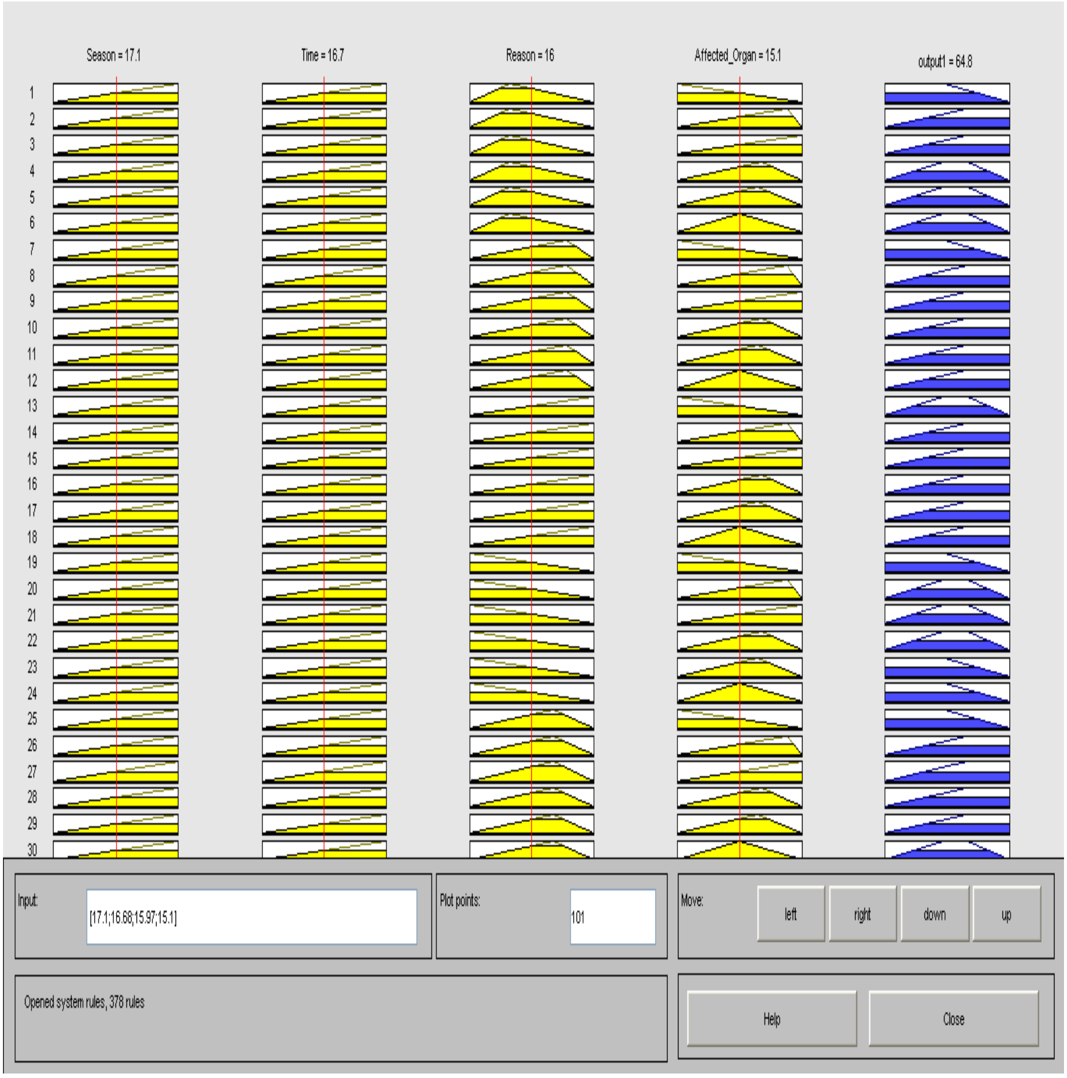


Figure 30. Graphical user interface for matlab's fuzzy rules for ELI

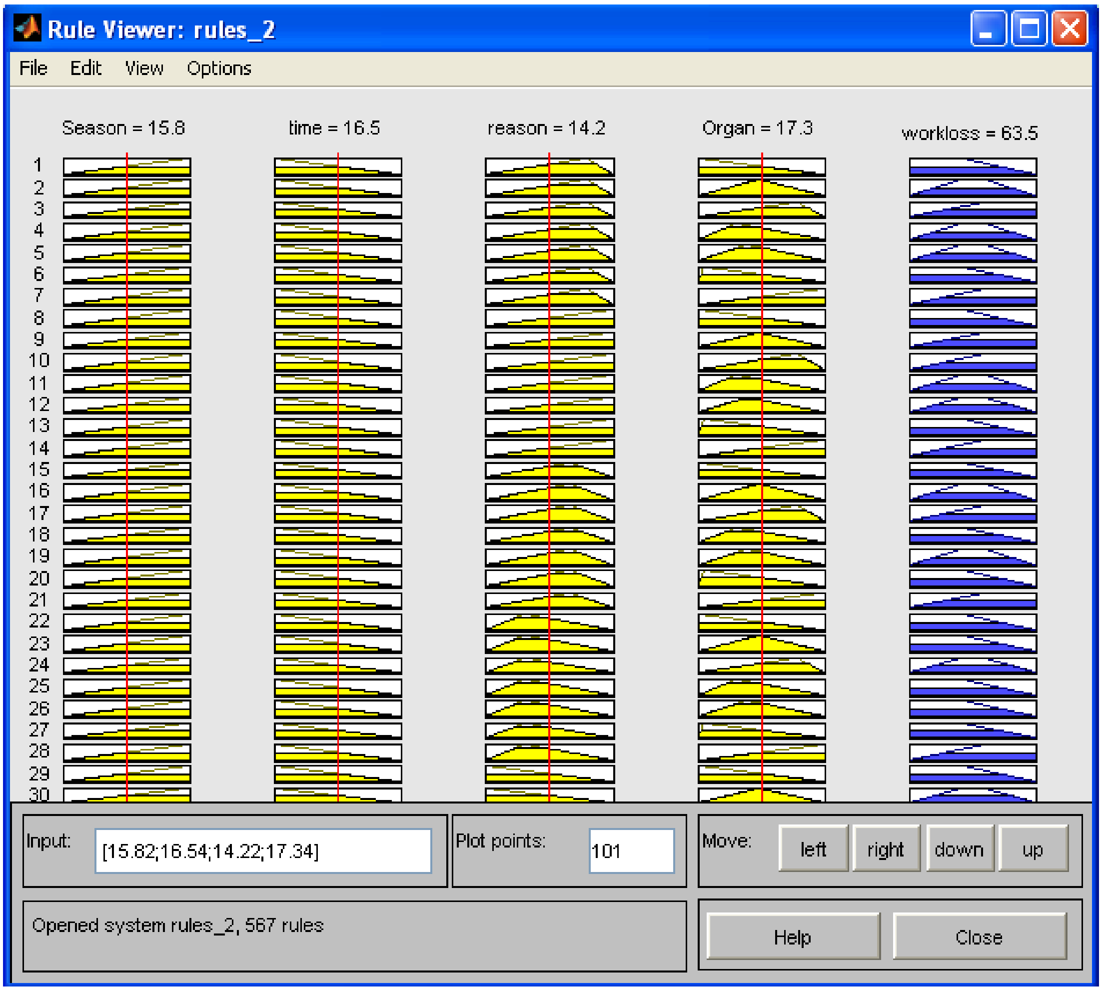


Figure 31. Graphical user interface for matlab's fuzzy rules for GLI

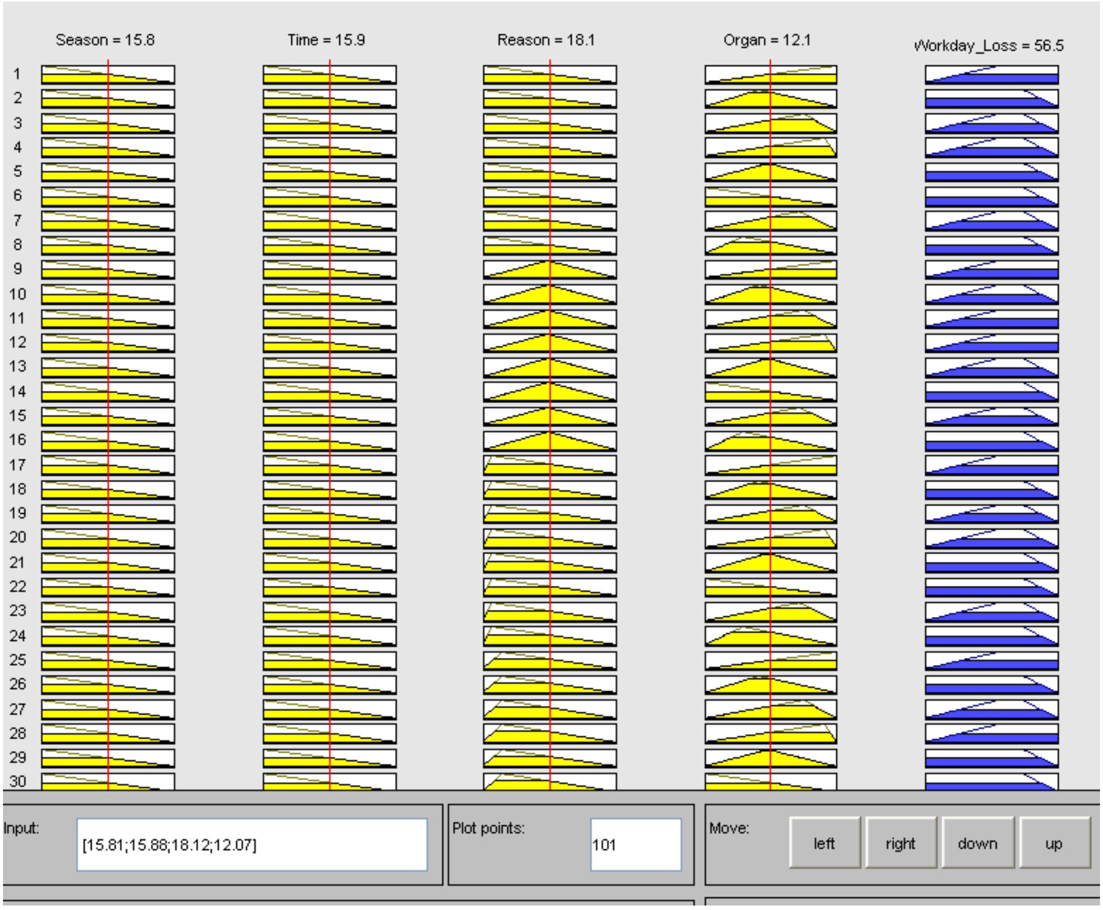


Figure 32. Graphical user interface for matlab's fuzzy rules for TTK

Figure 33 shows a surface for severity of accidents versus reason and affected organ as a sample, the other figures are given in Appendix 6. In the figure, two horizontal axes are (i) Affected organ and (ii) Reason, which represent independent variable. The result 'output1' is the resultant of four parameters, but Figure 33 is drawn assuming two of them as constant. Values on x-axis represent the partial effect of the two, while those on y-axis represent the overall value of output

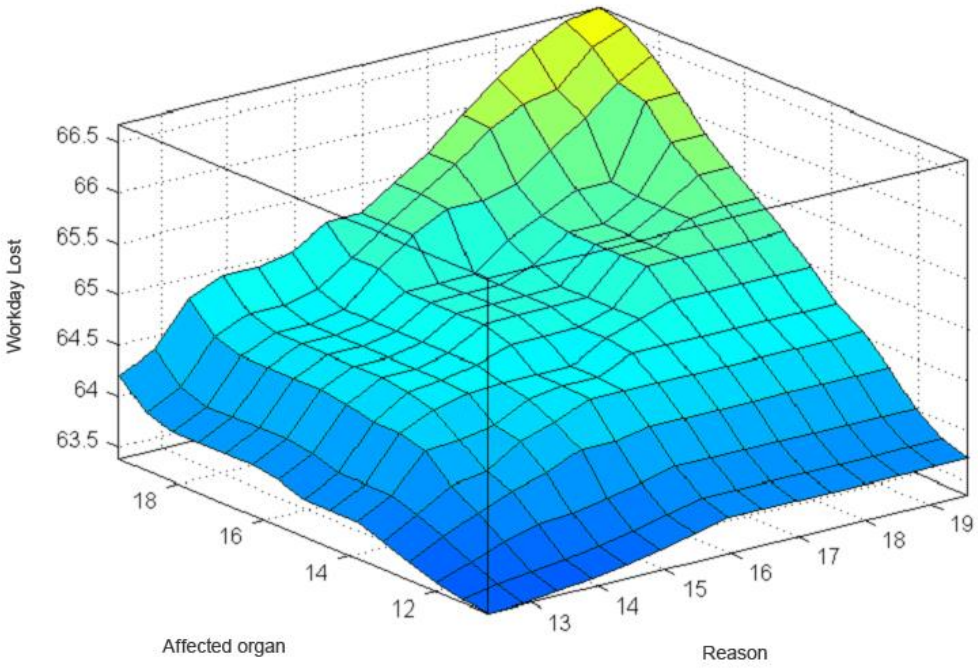


Figure 33. Surface for output1 (Workday Loss) for GLI versus reason and affected organ effect

Fuzzy logic estimator produces output from the given values using a set of rules, rather than the set of equations. Since the results of this method are based on quality attributes, they cannot be directly quantified for comparison like the other two methods. Also, fuzzy logic produces results 'expected average results' for combined efforts, instead of precise contribution of each variable/state. Since this is a different type of result, it cannot be compared with results obtained from regression.

However, it can be seen that most of the results are close to that predicted by the model. For instance, observation 11 for ELI shows the value of workday loss as '30' for (spring season, 0800-1600 shift, other injuries and hands), which is much higher than median (calculated as '9') value while that for observation 9 shows '6', which is much lower than median. Corresponding values predicted by fuzzy logic were also 3, 1. To a large extent, most of the findings from fuzzy logic matched for the entire picture. Also, this method proves that some results could be directly obtained from common sense, and were found to be similar to those obtained by regression. For instance, average value for data corresponding to ‘spring’ was found to be higher than for other seasons. This enables spring to have a relatively positive effect as compared to other seasons. Similarly, values for head injury were found to be lowest for both the sites. Other results found in Table 52 can also be verified by fuzzy logic method as well. In addition, there are many more results, such as reason traffic accidents (9) may also result in high workday loss. These results cannot be put as final results of the study because they could not be verified by either neural network or simple regression.

The method of replacement effect for fuzzy logic appears to have the same drawback as that for neural network – the values do not represent direct unit contributions from variables. Still, analysis of replacement effect can be done in the same way as that for regression. This is because a direct relation between output and sum of inputs has already been established under our assumptions. By doing the replacement analysis in the similar way as in regression, some good results were obtained. It was found that the following is the order of influence of parameters: affected organ, reason, shift and season. The latter two had opposite trend which is decreasing effect for both sites, but shift scored higher overall. This order is slightly different than that found by regression or neural networks analysis.

To summarize, it can be said that fuzzy logic produces good outputs for datasets having a large number of states, provided that the number of parent variables is not too large and the number of observations in dataset is sufficiently large. It is an improvement in this regard, over those methods whose accuracy deteriorates with increasing data size. However, it is not advisable for datasets having small number of discrete variable states. Codes for fuzzy logic may be found in Appendix 3.

CHAPTER 8

FUTURE ACCIDENT ESTIMATION

Another approach for estimating number of accidents related with the number of workers, number of working hours, total raw production and type of the coal (lignite or hard coal).

Sari 2002, used regression and time series modelling for projection of mine accidents and Regression and neural network models is used to have a comparison of the techniques and give a decision to choose the best model for the estimation of number of accidents between the study of Sari, 2002

In regression analysis; two regression statistics, significance level (*P* value) and coefficient of determination (R^2), were used for determination of variables to be eliminated. The *P* value gives an indication of the significance of the variables included in the model, whereas *R*2 gives a measure of the variability explained by the model. (Sonmez, 2004)

In neural network model for future accident estimation; feed-forward model is used. The input variables for the estimation of future accidents are; Number of Workers (W), Number of Working Hours (H), Total raw coal production (P) and type of coal (T) which are the independent variables of the models. The output variable is the number of accidents (Y).

MSE (Mean Standard Error) is used for the prediction performance of the models. Higher MSE gives the prediction performance poor.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{actual}_i - \mathrm{predicted}_i)^2 \tag{8.1}$$

8.1 ELI Models

Table 53 gives information about yearly number of accidents, number of deaths, number of injured, total working hours, total raw coal production and type of coal mine. It can be seen that while total raw coal production and working hours decrease, number of accidents are decreased.

Table 53. early accident and production in ELI mine

Year	Accidents	Death	Injured	Total Worker	Number of Working hours	Total raw coal production (ton)
2011	29	1	28	1675	3486312	5424326
2010	24	1	23	1816	2917520	5424326
2009	33	1	32	1981	4169376	6204952
2008	35	0	35	1978	4380672	6204952
2007	47	1	46	2429	4881072	5708914
2006	90	0	90	3053	5619296	5467611
2005	180	1	179	2963	6650016	6570953
2004	275	0	275	3066	6379400	6849597
2003	256	0	256	2963	7195776	6849597
2002	258	0	258	3552	7944416	6849597
2001	317	0	317	3718	8260656	9436845
2000	355	0	355	3987	8444088	11131520
1999	405	4	401	4039	8181280	11105167
1998	583	2	581	4134	8918280	10868230
1997	491	2	489	4357	8778344	11335959

Regression Analysis

$Y = \beta_0 + \beta_1 W + \beta_2 H + \beta_3 P$

Where;

β_x = Constants

W = Number of Workers

H = Number of Working Hours

P = Total raw coal production

When the analysis is applied for ELI data including all input variables; the equation is;

$Y = -376{,}342 + 0.044\ W + 3.6\ 10^{-5}\ H + 2.95\ 10^{-5}\ P$

Significance level is chosen as 95%.

Table 54. Elimination of Irrelevant Input Variables in ELI mine

Model	Independent Variables	R^2	Variable corresponding to the coefficient with the highest P value	Significance Level (P value)	MSE
1	W, H, P	0.903	W	0.635	61.927
2	H,P	0.901	9	0.028	59.671

After calculation of P values which indicates lower P value means the equation is more significant.

it is understood that number of workers will not have a significant effect on the model therefore it is excluded from the model and the final model is achieved with 2 input variables which are number of working hours and total raw coal production.

The final equation is;

$Y = -365.987 + 5.21.10^{-5}\ H + 3.31.\ 10^{-5}\ P$

Table 55. P Values of the Model in ELI

Independent Variable	Significance Level (P value)
Number of Working Hours (H)	0.007
Total Raw Coal Production (P)	0.028

Neural Network Analysis

Table 56. Neural Network Models for ELI Mine

Model	Independent Variables	Number of Hidden Neurons	R^2	MSE
1	W, H, P	5	0.83481	13776
2	W, H, P	10	0.922	357.0407
3	W, H, P	15	0.904	14.11
4	W, H, P	20	0.951	15.71
5	H,P	5	0.979	1.609
6	H,P	10	0.987	0.611
7	H,P	15	0.964	0.025
8	H,P	20	0.984	1.626

The neural network model is performed using 3 input variables (number of workers, number of working hours and total coal production) and accident number as an output and analysed with different number of hidden neurons, for the best prediction it is also performed 2 variables, the most irrelevant from regression analyses is excluded from the model.

Similar with regression analysis final model is chosen with 2 variables (number of working hours and total raw coal production) with 10 neurons which gives best fit (R^2 = 0.987) and lowest mean standard error.

8.2. GLI Models

Table 57. Yearly accident and production in GLI mine

Year	Accidents	Death	Injured	Total Worker	Number of Working hours	Total raw coal production (ton)
2011	23	0	23	1739	3715144	5696000
2010	15	0	15	1756	2827224	3814406
2009	32	0	32	1900	4455832	3814406
2008	72	0	72	2051	8145480	3814406
2007	58	1	57	2505	5215800	2988772
2006	48	1	47	2585	5500544	3449129
2005	85	0	85	2815	5893088	3324506
2004	96	0	96	3209	6638680	3734290
2003	111	0	111	3402	6960040	3734290
2002	152	1	151	3779	7608928	3734290
2001	175	0	175	3777	8004032	3958682
2000	157	1	156	4077	8276592	4166956
1999	188	2	186	4162	8815408	3262574
1998	190	0	190	4236	9008584	3874103
1997	231	0	231	4426	8348440	4103649

Regression Analysis

$Y = \beta_0 + \beta_1 W + \beta_2 H + \beta_3 P$

Where;

β_x = Constants

W = Number of Workers

H = Number of Working Hours

P = Total raw coal production

When we apply the analysis for GLI data including all input variables; the equation is;

$Y = -156.549 + 0.055\ W + 9.02\ 10^{-6}\ H + 9.5\ 10^{-6}\ P$

Significance level is chosen as 95%.

Table 58. Elimination of irrelevant input variables in GLI mine

Model	Independent Variables	R^2	Variable corresponding to the coefficient with the highest P value	Significance Level (P value)	MSE
1	W, H, P	0.953	P	0.23	16.86
2	W,H	0.946	H	0.084	17.25

The final equation is;

$Y = -115.994 + 0.05\ W + 8.47.\ 10^{-6}\ H$

Similarly with ELI model P values are calculated and for GLI, number of working hours and number of workers chosen for the model.

Table 59. P Values of the model in GLI

Independent Variable	Significance Level (P value)
Number of Workers (W)	$6.18.10^{-5}$
Number of Working Hours (H)	0.084

Neural Network Model

Table 60. Neural network models for GLI mine

Model	Independent Variables	Number of Hidden Neurons	R^2	MSE
1	W, H, P	5	0.928	1075.48
2	W, H, P	10	0.952	158.12
3	W, H, P	15	0.978	6756.7811
4	W, H, P	20	0.958	0.495
5	H,P	5	0.997	33.26
6	H,P	10	0.982	341.15
7	H,P	15	0.995	50.97
8	H,P	20	0.999	39.77

The neural network model is performed using 3 input variables (number of workers, number of working hours and total coal production) and accident number as an output and analysed with different number of hidden neurons, for the best prediction it is also performed 2 variables, the most irrelevant from regression analyses is excluded from the model.

Similar with regression analysis final model is chosen with 2 variables (number of working hours and number of workers) with 5 neurons which gives best fit (R^2 = 0.997) and lowest mean standard error.

8.3 TTK Models

Table 61. Yearly accident and production in TTK mine

Year	Accidents	Death	Injured	Total Worker	Number of Working hours	Total raw coal production (ton)
2011	2809	4	2805	11104	24484320	1592515
2010	3478	5	3473	11456	25260480	1708844
2009	3555	7	3548	10979	24208695	1879630
2008	1526	5	1521	9685	21355425	1586532
2007	2074	5	2069	10553	23269365	1675373
2006	1679	3	1676	10611	23397255	1522421
2005	1850	10	1840	11249	24804045	1665324
2004	2220	4	2216	12261	27035505	1879411
2003	2488	7	2481	14062	31006710	2011178
2002	2664	7	2657	15761	34753005	2244372
2001	4232	1	4231	18025	39745125	2356865
2000	4037	1	4036	19151	42227955	2259277

Regression Analysis

$Y = \beta_0 + \beta_1 W + \beta_2 H + \beta_3 P$

Where;

β_x = Constants

W = Number of Workers

H = Number of Working Hours

P = Total raw coal production

When the analysis is applied for TTK data including all input variables; the equation is;

$Y = -752.209 + W + 4.9\ 10^{-5}\ H + 0.0011\ P$

Significance level is chosen as 95%.

Table 62. Elimination of irrelevant input variables in TTK mine

Model	Independent Variables	R^2	Variable corresponding to the coefficient with the highest P value	Significance Level (P value)	MSE
1	W, H, P	0.505	P	0.606	714.28
2	W,H	0.490	-	-	688.23

According to P values, number of working hours as an input is excluded from the model.

The final equation is;

$Y = 68.19 + W + 9.31.\ 10^{-5}\ H$

Table 63. P Values of the model in TTK

Independent Variable	Significance Level (P value)
Number of Workers (W)	≅ 0
Number of Working Hours (H)	≅ 0

Neural Network Model

Table 64. Neural network models for TTK Mine

Model	Independent Variables	Number of Hidden Neurons	R^2	MSE
1	W, H, P	5	0.791	240.617
2	W, H, P	10	0.869	255646
3	W, H, P	15	0.912	92742
4	W, H, P	20	0.955	4683
5	H,P	5	0.621	77620
6	H,P	10	0.987	533818
7	H,P	15	0.978	686790
8	H,P	20	0.990	2421455

The neural network model is performed using 3 input variables (number of workers, number of working hours and total coal production) and accident number as an output and analysed with different number of hidden neurons, for the best prediction it is also performed 2 variables, the most irrelevant from regression analyses is excluded from the model.

Similar with regression analysis final model is chosen with 2 variables (number of working hours and number of workers) with 10 neurons which gives best fit (R^2 = 0.997) and lowest mean standard error. TTK models the MSE is much higher than the other mines, due to the high number of data is used.

CHAPTER 9

RESULTS AND DISCUSSIONS

9.1. Introduction

In this part of the study the results will be presented and comparative analysis between the findings will be made.

In the research a risk analysis study was conducted for ELI, GLI and TTK mines. After risk analysis; according to workday loses, accident severity estimation models were developed using three different methodologies which are regression, neural network and fuzzy logic. Future estimation models were also developed using regression analysis and neural network with number of total workers, working days and production.

Different methods were used for analysis of dataset, because each one has its own set of advantages and disadvantages against the others. Also, they could be used to validate the results of one another, or overcome the drawbacks of other methods to generate some results which could not be detected by other methods.

Apart from the challenge of comparing the effectiveness of various methods available for analysis, another challenge arises from interpretation of results. The results are in the form of ‘the factor A has a higher workday loss per accident, than B’.

A comparative analysis for three mines was performed in the current work, and was useful in determining the course of action, in case states of categorical variable were to be distributed in various sites.

9.2 Comparison of Researches

Sarı (2002) conducted a risk modelling study for ELI and GLI mines using a 5 year period data. The main aim of the study was to develop main risks in lignite mines using risk matrix methodology with probability and severity variables. He also made a future estimation with regression and time series modelling.

Considering as a starting point of the study of Sarı (2002), the current study is expanded using 10 year period data of ELI and GLI mines and including a 4 year data of TTK mines which produces hard coal.

Risk assessment methodology was also developed with Fine and Kinney (1976) methodology using frequency variable in the risk equation. This method has an advantage comparing risk matrix methodology as it divides the failure rate or frequency of occurrence into two factors. The method gives one more aspect to consider.

When a comparison made between the risk analyses; Sarı (2002) found that manual handling and falling rocks has highest risk and struck by falling objects, haulage, slip/falls and hand tools follows respectively and the most effected organs were found as feet, main body and hands in ELI mine. In the current study; manual handling is found as a high risk too but machinery is also another high risk reasons, the most effected organs are same with the study of Sarı (2002) the only difference is the sequence of high risk organs as hands, feet and main body. The risk analysis is not so different in GLI mines, in the research of Sarı (2002) manual handling and struck by falling object has highest risks with the most effected organs as feet and main body, in the present study falling rocks has also a high risk in GLI mines with including hands affecting organ.

Another comparison can be made with the probability component of the studies. Sarı (2002) found that the accident has a probability of 0.4 per day in GLI mine and 0.72 in ELI mine which has changed in the current study as 0.32 in GLI mine and 0.66 in ELI mine. This change can be the result of drop of production numbers and decrease in the number of workers.

9.3. Comparison of Regression, Neural Network and Fuzzy Logic Techniques

The current analysis was performed with regression in two forms. Simple regression was found to be weak for this study, and modified regression with parsimonious approach was found to be even weaker, and did not produce higher correlation for any of the TKI and TTK mines. However, one good thing which came up from the regression part is a finding - which the method does not work in presence of a large number of categorical variables. Similarly, it can be said that parsimonious regression model may not produce results better than simple regression in all conditions. Even though it did produce some results, it was difficult to say whether the relations were correct or co-accidental. Most of them were verified by other methods as well, but a few were found to be different/opposite. Neural networks method could produce new results, by improving the efficiency of simple regression method, by 'training' the entire data, instead of removing a few inputs. This approach worked and produced significantly better outputs. Fuzzy logic provided even better results.

Regression analysis is the simplest and easiest to understand and implement, and could be performed by just converting the data to a suitable form, and applying a least squares fit line to it. Thus, it is relatively quick & simple for machine to process, because it produces results in a single step matrix computation. It is also easy for humans to implement, because once the discrete variables of the datasets are converted to the desired format, the regress function directly produces the results. For regression, results will be same for a given dataset, irrespective of the user or software performing it. This is because regression values are precisely defined by a straight mathematical expression, and this is not true for neural networks or fuzzy logic method. However, a big disadvantage of this simple regression method is its poor accuracy in comparison to other two methods for samples with large number of variables (or discrete variable states). It is good enough for samples having small number of variables, but performs poorly if the number of states plus continuous variables is large.

The problems with regression get reduced to a large extent by using neural networks. However, this method involves several steps, such as creation of training interface, iterative translation of output data for improvement of coefficients, and finally regression as well. Also, it produces different results for the same dataset. The following are the reasons of such variations

- Variation in number of neurons
- Percentage of points allocated to training, validation and testing set
- Selection of initial random seed

Even if all conditions are fixed, the method produces different results on different runs because of the fact that the value of the initial random 'seed' selected by the method decided the values of random observations selected by the system at each step. Problems due to these factors could be reduced to a considerable extent by doing the following

- If results are provided with number of neurons as one of the parameters, they can be used as per the neuron requirements for further works – this is another reason why number of neurons was kept as a user-defined parameter
- If the percentages of data allocated for training, validation and testing sets are kept same for all work (such as the default values of the software), some standardization might be achieved - for example, Matlab standards of 0.7, 0.15, 0.15 were taken as it is, in this work

- If seed value is fixed by the programmer before system picks a random seed, the results of two consecutive runs will not vary.

As already explained, another problem with this method is that it carries most of the drawbacks of its parent method – regression, because after the processing of data, it also uses regression to compute final values. However, values are better because inputs and output values are modified before regression.

Third method 'fuzzy logic', removes some of the disadvantages from regression, and neural networks method because it provides a better quality solution without requiring complex iterative methods. However, on the downside, it requires lots of human skills, time and efforts for development of rules. Usually, fuzzy rules are generated by applying some algorithm to the existing data. Such an algorithm varies from problem to problem. Also different developers may come up with different approaches because of which, in general, results will also be different. Though machine implements the rules exactly in the way expected by the developer, and there are no chances of variation on its part, but this variation due to human counters the machine variation for neural networks method. Moreover, unlike for neural networks method, this is a source of inconsistency which can neither be modified nor eliminated. Also, the method is further time consuming because as per that, one set of rules cannot be applied for a different set of problem. This problem can be partially overcome by implementing a generalized algorithm through a Matlab code – the one which reads and processes datasets to generate the rules as long as basic format is similar. The code developed for this study was designed for only four columns, since both the sites had only four discrete variables (season, affected organ, shift and reason of accident), but it could work for two sites where the number of 'significant states' for variables were different, and hence the number of rules also. The time and effort required for feeding the rules to fuzzy logic toolbox could still not be eliminated.

As a summary, simple regression method scores positive human toughness, machine complexity, result variation by human/machine. However, intuition suggests that a method which produces poor results is not desirable even if it is very simple and precise. In other words, the objective of data analysis is to set up a compromise between result accuracy and other parameters. Considering all this, it can be said that fuzzy logic is better than the other two methods for analysis of the type of data given for ELI and GLI. On the other hand, different method could be used if data set were of a different type. For instance, if dataset had no discrete variables, neural networks, or even regression, could have worked better than fuzzy logic, and even produced better results. On the other hand, if dataset were even more qualitative, fuzzy logic would have produced even better results.

Although fuzzy logic has a very complex structure and difficult to set the rules; it gave very good results for the estimation of workday loss.

CHAPTER 10

CONCLUSIONS AND RECOMMENDATIONS

Risk estimation and analysis techniques are very crucial for managing underground coal mine risks due to many fatal accidents and equipment loses which cost millions of liras. This study can help mines to ensure their safety with productivity using past accident data and can give an estimate for the future.

The reasons of accidents and in more detail the relationship between reasons and causes of accidents were analysed using previous accident data covering 14 years of TKI and 4 years for TTK mines. Current accident data includes workday losses, time, season, reason of the accident and age, affected organ of the injured.

The main conclusions and recommendations can be summarized;

- The findings of the study provide an understanding to use neural network and fuzzy logic techniques for estimation of severity
- Manual handling is a common risk in all coal mines
- Hard coal mines are much more hazardous than lignite mines considering accidents per day. (The probability of accident is in TTK mines is 18.28 times higher in GLI and 8.86 times higher in ELI)
- Hard coal mines have different risk reasons compared to lignite mines. Fatalities are dominantly related to gas/dust explosions, blasting and strata problems in hard coal mines while machinery is the main risk in lignite mines.

- Total raw coal production and number of working hours has a direct impact on accident severity which has approximately 0.90 regression value in all models.
- The effect of variables on accidents can be summarized as;

  Age: The effect of the only continuous variable, age, was not found to be significant enough based on any of the three methods. It could not be established whether increasing age will have an increasing effect or reverse. So, age factor can be considered to have a weak or negligible effect. Nevertheless, it is true that age was confined in a very narrow range in the given dataset. The effects of age could not be measured because of this reason.

  Season: Spring found to have significantly higher adverse effect in all methods when compared to other seasons. However, other three seasons did not have a stronger replacement effect. Overall, season effect was weak, but spring is responsible for enhancing the severity of accidents.

  Time: Shift was found to have average effect – first shift (0800-1600) was detected as responsible for enhancing the severity of accidents, by neural networks method, and was later verified by fuzzy logic. However, overall effect of time was found to be weaker than that for season.

  Reason: This was a strong influencer, having maximum 14 states. State 7 (Manual Handling) was found to have significantly higher than the median value for lignite mines where it had highest value. Thus, it can be claimed that manual handling is definitely responsible for enhancing the severity of incidences. In lignite mines, falling rocks, equipment related accidents, machinery follows respectively. In hard coal mine blasting and gas/dust explosions has enhancing effect followed by struck by/against object and manual handling.

  Affected Organ: Affected organ was another strong influencer, which had a maximum of 8 states. Several results were obtained from this set. They are (i) head (1) injuries have reducing effect, and (ii) foot and arm (4) injuries had enhancing effects on severity of accidents. Relatively, the effects of (1, 3) were stronger than that for (4). This completes (i) the understanding of the manner in which various causes are related to accidents, and (ii) the effect of various states of the given variables (causes) on accidents.

Near miss reporting data were absent. As it is known that near miss reporting is a very important tool for accident investigations in a deep extent to prevent accidents. In addition to near miss reporting, there were no accidents data related to damage to the equipment or machinery.

It is very important to develop a model and make a comparison between each other; the data should be in the same form, instead of different types of data collection forms in different mines. In the study, a lot effort consumed to make the data in compatible to each other using pre-processing. The mines collect the data and must form a database including near miss accidents.

The study can be extended to other underground mines by considering the production methodologies like conventional or mechanized and comparison between them.

Qualitative risk assessments methodologies specific to each mine must be implemented. Hence, a further study can deal with one workplace and applying techniques including the judgement of the expert. This kind of study can give much more detailed results and can make a comparison between quantitative methodologies.

REFERENCES

1. Arda M., (2008) *Risk Assessment for a denim manufacturing plant in Turkey.* Middle East Technical University

2. Bagot, A. (2009). *Accidents in Mines: Their Causes and Prevention.* Retrieved from books.google.com

3. Bingham, N.H. and Fry, J.M. (2010). *Regression: Linear Models in Statistics*. London: Springer

4. Broadleaf Capital International Pty Ltd (BCI).(2007).*Tutorial Notes: The Australian and New Zealand Standards on Risk Management*. Retrieved from www.Broadleaf.com.au

5. Bu-Quammaz A.S. (2007) *Risk Assessment of International Construction Projects Using the Analytic Network Process.* Ankara: Middle East Technical University. Department of Civil Engineering

6. Chong M.M., Abraham A., Paprzycki M. (2004) *Traffic Accident Analysis Using Decision Trees and Neural Networks*. USA. Computer Science Department, Oklahoma State University.

7. Epp, D. J. et. al.,(1977). *Identification and specification of inputs for benefit-cost modeling of pesticide use, Volume 1*. Washington D.C. : Environment Protection Agency, Office of Research and Development, Office of Air, Land and Water use

8. Ericson, C.A., (2005). *Hazard Analysis Techniques for System Safety*. Virginia: John Wiley and Sons Inc.

9. Fesak, G.M., Cavanaugh, D. R. and United States Office of the Administrator, Coal Mine Safety and Health. (1985). Report of investigation: underground coal mine explosion, Greenwich Collieries no. 1 mine, I.D. no. 36-02405 : Greenwich Collieries Division of Pennsylvania Mines Corporation, Green Township, Indiana County Pennsylvania, February 16, 1984. Pennsylviana: U.S. Dept. of Labor, Mine Safety and Health Administration, Office of the Administrator, Coal Mine Safety and Health

10. Frankel, M., Hommel, U. and Rudolf, M. (2005).*Risk Management: Challenge and Opportunity (2nd edition)*. Berlin: Springe

11. Haight F.A. (1967). Handbook of the Poisson Distribution. New York: John Wiley & Sons.

12. Goodman D. Yates R.D. (2004) *Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers*.

13. Gözüyilmaz, C. (2003). A system approach to occupational health and safety management. Unpublished Manuscript

14. Gurney, K. and Gurney, N. K. (1997). *Neural Network.* London: CRC Press

15. Gurcanli G.E, Kuruoglu M., Mungen U. 2005, *An approach for hazard assessment on construction sites using fuzzy sets*. , Ergonomics, 39, 371-387.

16. Güyagüler T., September 2000, Occupational Health and Safety, Middle East Technical University, Lecture Book.

17. Han S. (2005) *Estimation of Cost Overrun Risk in International Projects By Using Fuzzy Set Theory*. Ankara: Middle East Technical University. Department of Civil Engineering

18. Hardy, M. A. (1993) .*Regression with Dummy Variables*, Issue 93. London: SAGE

19. Hashemi R.R., Blanc L., Rucks C., Shearry A. (1995). *A neural network for transportation safety modelling*. Expert Systems with Applications Vol 9., No.3 pp 247-256.

20. Heizer, J. , Rajashekhar, J. and Render, B.(2009). Operations Management(9th ed.). Noida: Pearson Education

21. Hong P., Lei-Jun L., Yun-jie L., *Model of Gas Warning System Based on Neural Network Expert*. 201O International Conference On Computer Design And Appliations.

22. Iannacchione, A. T., Brady, T. M. and Varley,F.(2008). The Application of Major Hazard Risk Assessment (MHRA) to Eliminate Multiple Fatality Occurrences in the U.S. Minerals Industry. Washington: U.S. Department of Health and Human Services

23. Karacan Ö. (2007) *Development and application of reservoir models and artificial neural networks for optimizing ventilation air requirements in development mining of coal seams.* International Journal of Coal Geology 72 (2007) 221–239

24. Karadas E. (2007) *A Risk Management Method for Turkish Defence Industry Firm.* Ankara: Middle East Technical University. Department of Business Administration.

25. Karmis, M.(Ed.).(2001). *Mine Health & Safety Management*. Colorado : Society for Mining, Metallurgy and Exploration Inc.

26. Kinney G.F., Wiruth A.D. Practical Risk Analysis for Safety Management, Naval Weapons Center, 1976

27. Langton, J. (1979). *Geographical Change and Industrial Revolution*. New York: Press Syndicate of the University of Cambridge

28. Lilic N., Obradovic I., Cvjetic A. *An intelligent hybrid system for surface coal mine safety analysis.* Engineering Applications of Artificial Intelligence 23 (2010) 453–462

29. Lind, D. A., Marchal, W.G. and Wathen S.A. (2005). *Statistical techniques in business & economics*. New Delhi: Tata McGraw Hill Education Private Limited

30. Masi, S., Mancini, I. M., Caniani, D., and Lioi, D. S. (2012): *Fuzzy logic and neuro-fuzzy networks for environmental risk assessment*, Open Access book "Fuzzy Logic".

31. Marhavilas P.K., Koulouriotis D., Gemeni V. Risk analysis and assessment methodologies in the work sites: On a review, classification and comparative study of the scientific literature of the period 2000-2009. (2011), Journal of Loss Prevention in the Process Industries 24 (2011) 477-523

32. Marhavilas P.K. Risk Estimation in the Greek Constructions Worksites by using a Quantitavive Assessment Technique and Statistical Information of Occupational Accidents, Journal of Engineering Science and Technology Review 2(2009) 51-55.

33. McDonald L.B., Pomroy W.H. and United States Bureau of Mines (1980). *A statistical analysis of coal mine fire accidents in the United States from 1950 to 1977*.US Department of the Interior, Bureau of Mines,8830,1-42

34. Moghaddam F.R., Afandizadeh S., Ziyadi M., Prediction of accident severity using artificial neural networks,International Journal of Civil Engineering, 2010.

35. Mustakim F., Yusof I., Rahman I., Samad A. Salleh N.(2008) *Blackspot Study and Accident Prediction Model Using Multiple Liner Regression.* Karachi, Pakistan. First International Conference on Construction In Developing Countries

36. NIOSH Mining Safety and Health Research Program's review committee, Committee on Earth Resources, National Research Council.(2007). *Mining Safety and Health Research at NIOSH: Reviews of Research Programs of the National Institute for Occupational Safety and Health*. Washington DC: National Academies Press

37. NSW Coal Association (NSWCA), Training & Safety Committee, New South Wales Coal Association. Training and Safety Committee. (1990). *Electrical Hazard Awareness for Surface Mines.* Retrieved from books.google.com

38. Occupation Safety and Health Administration (OSHA), US Department of Labor. (2002). *Job Hazard Analysis*. Retrieved from http://www.osha.gov/Publications/osha3071.pdf

39. Oleson, J.P., (Ed.).(2008). The Oxford Handbook of Engineering and Technology in the Classical World. Oxford: Oxford University Press

40. Oxford Dictionaries.(2011). *Concise Oxford English Dictionary: Main edition*. Oxford: Oxford OUP

41. Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers, Roy D. Yates, David Goodman, page 60.

42. Radosavljević S and Radosavljević M. (2009) *Risk Assessment in Mining Industry: Apply Management.* Serbian Journal of Management 4 (1) (2009) 91 – 104

43. Ross, T. J. (2009). *Fuzzy Logic with Engineering Applications*. Chichester: John Wiley & Sons

44. Richards, J. S., (2007). Death in the Mines: Disasters and Rescues in the Anthracite Coal Fields of Pennsylvania. Charleston: The History Press

45. Ruilin Z., Lowndes I., *The application of a coupled artificial neural network and fault tree analysis model to predict coal and gas outbursts.* International Journal of Coal Geology 84 (2010) 141–152

46. Sando T. (2005) *Modelling Highway Crashes Using Bayesian Belief Networks Technique and GIS*. Florida State University College of Engineering.

47. Sari, M. (2002). *Risk Assessment Approach on Underground Coal Mine Safety Analysis*. Ankara: Middle East Technical University. Department of Mining Engineering

48. Sari M., Karpuz C., Selcuk A.S., An International Comparison of Turkish Coal Mining Industry Safety Performance, Mine Planning and Equipment Selection 2001.

49. Sebnem, A.(2009). *Mine Blast Kills 19 in Turkey*. New York Times.pp.1

50. Sii H.S., Wang J., Ruxton T., Yang J.B., Liu J. (2004). Fuzzy logic approaches to safety assessment in maritime engineering applications, Journal of Marine Engineering & Technology, IMarEST

51. Singhal, R.K. and Malhotra, A. K.(2000). Environmental Issues and Management of Waste in Energy and Mineral Production: Proceedings of the Sixth International Conference on Environmental Issues and Management of Waste in Energy and Mineral Production, SWEMP 2000, Calgary, Alberta, Canada, May 30-June 2, 2000.Netherlands: Taylor and Francis

52. Stellman, J. (1998). Encyclopaedia of Occupational Health and Safety: Industries and occupations, Volume 3 (4^{th} ed.). Geneva: International Labour Organization

53. Sonmez R. Conceptual cost estimation of building projects with regression analysis and neural networks. NRC Canada, 2004.

54. Yang L., Song M., *Coal Mine Safety Evaluation with V-Fold Cross-Validation and BP Neural Network*. Journal of Computers Vol. 5, No.9 September 2010.

55. Y. Yuan and M.J. Shaw, Induction of fuzzy decision trees. Fuzzy Sets and Systems 69 (1995), pp. 125–13

APPENDIX A.

CODES FOR NORMAL REGRESSION

6.1 main file (master.m)

```
function pfm = master (inputfile,outputfile)
    % x contains independent parameters' columns
    % y is the response column
    [x,y,c] = xlinput (inputfile);
    % all values have 2 states - on/off represented by their...
    x1 = [x(:,2:5)];
    x2 = [x(:,1)];
    x3 = dummyvar(x1);
    dx = [x2 x3 ones(size(x3(:,1)))];
    [b,bint,r,rint,stats] = regress (y,dx); % regresses dummy matrix against workday
loss  o = xloutput (outputfile,b,stats(1),c);
```

6.2 xlinput.m

```
function [x1,y1,c] = xlinput(file)
x1 = [];
y1 = [];
c = [];
d1 = xlsread(file, 1, 'A:A');
%f1 = findmax(d1');
c = [c 1]; % first column is continuous variable
x1 = [x1 d1];

d2 = xlsread(file, 1, 'B:B');
```

```
f2 = findmax(d2')+1;
c = [c f2];
x1 = [x1 d2];
d3 = xlsread(file, 1, 'C:C');
f3 = findmax(d3')+f2;
c = [c f3];
x1 = [x1 d3];
d4 = xlsread(file, 1, 'D:D');
f4 = findmax(d4')+f3;
c = [c f4];
x1 = [x1 d4];
d5 = xlsread(file, 1, 'E:E');
f5 = findmax(d5')+f4;
c = [c f5 f5+1]; %column for constant
x1 = [x1 d5];
y1 = [xlsread(file, 1, 'F:F')];
```

6.3: xloutput

```
function z = xloutput (file,m,r,c)
a1 = m(1:c(1));
a2 = m(c(1)+1:c(2));
a3 = m(c(2)+1:c(3));
a4 = m(c(3)+1:c(4));
a5 = m(c(4)+1:c(5));
a6 = m(c(6));
xlswrite(file,a1,'Age');
xlswrite(file,a2,'Season');
xlswrite(file,a3,'Time');
xlswrite(file,a4,'Reason');
xlswrite(file,a5,'Affected Organ');
xlswrite(file,a6,'Constant');
xlswrite(file,r,'R^2');z = 0
```

APPENDIX B

CODES FOR MODIFIED REGRESSION

7.1 main file (master.m)

```
function pfm = master (inputfile,outputfile)
   % x contains independent parameters' columns
   % y is the response column
   [x,y,c] = xlinput (inputfile);
   % all values have 2 states - on/off represented by their...
   % corresponding dummy variables
   x11 = [x(:,2:5)];
   x12 = [x(:,1)];
   x13 = dummyvar(x11);
   %dx = [x12 x13 ones(size(x13(:,1)))];
   dx = [x12 x13];
   ncol = size (dx,2);
   [ir,xf] = pregress (dx,y);
   [m,bint,r,rint,stats] = regress (y,xf); % regresses remaining dummy matrix against
workday loss
   o = xloutput (outputfile,m,stats(1),c, ir,ncol);
```

7.2 pregress.m

```
function [ir,xf] = pregress (dx,y)
%[b,dev,stat] = glmfit (x,y,'normal','constant','off'); % x already contains ones
[b,dev,stat] = glmfit (dx,y,'normal','constant','off');
```

```
p = stat.p;
ir = find(isnan(p)); % create array for NaN elements
ir = sort(ir');
dx(:,ir) = []; %remove columns from x containing NaN elements
p(ir) = []; % same for p
p2 = p;
dx2 = dx;
nv = [];
[pmax,pmaxind] = max(p2);
p20=p2';
j1 = 0;
while (pmax>0.2 && size(dx2,2)>2)
   dx2(:,pmaxind) = []; % remove column corresponding to pmaxind (dynamix)

   for i=1:numel(ir)
      i1 = ir(i);
      p20 = [p20(1:i1-1) 0 p20(i1:numel(p20))];
   end
   % this matrix contains zeros for non-existing locations
   [pmax0, pmaxind0] = max(p20); % this index value will be appended to ir matrix
   nt = numel(ir);
   %ne = numel (ir(ir<pmaxind0)); % number of elements before index value to be
removed
   % ir = [ir pmaxind0 find(isnan(p20))]; % accounting for na
   ir = [ir pmaxind0];
   %if (ne<pmaxind0)
    %   ir = [ir ir(ne+1:nt)]; % inserting
   %end
   ir = sort(ir); %sorting matrix

   %[b,dev,stat] = glmfit (dx2,y,'normal'); % fitting again
```

```
    [b,dev,stat] = glmfit (dx2,y,'normal','constant','off');
    p2 = stat.p; % new value of p2
    p20=p2';
    [pmax,pmaxind] = max(p2); % computation of maximum for next verification
    p2
    ir
end

xf = dx2;
```

7.3 xloutput.m

```
function z = xloutput (file,m,r,c,ir,ncol)
m1 = 1:ncol;
ins = m1;
ins(ir) = [];
m1 = zeros (30,1);
m1(ins) = m';

a1 = m1(1:c(1));
a2 = m1(c(1)+1:c(2));
a3 = m1(c(2)+1:c(3));
a4 = m1(c(3)+1:c(4));
a5 = m1(c(4)+1:c(5));
a6 = m1(c(6));
%a6 = m(c(5)+1:c(6));
%a7 = m(c(7));
xlswrite(file,a1,'Age');
xlswrite(file,a2,'Season');
xlswrite(file,a3,'Time');
xlswrite(file,a4,'Reason');
xlswrite(file,a5,'Affected Organ');
```

```
%xlswrite(file,a6,'Organ');
xlswrite(file,a6,'Constant');
xlswrite(file,r,'R^2');
z = 0;
```

7.4 xlinput.m (same as 6.2)

APPENDIX C

CODES FOR NEURAL NETWORKS

8.1 main file (master.m)

```
function pfm = master (inputfile,nfile,outfile0,outfile,xfile,yfile)
    % x contains independent parameters' columns
    % y is the response column
    [x,y,c] = xlinput (inputfile);
    % n is the number of layers
    nmatrix = csvread (nfile); % matrix of n
    n_e = numel(nmatrix); % no. of elements in n
    x1 = [x(:,2:5)];
    x2 = [x(:,1)];
    x3 = dummyvar(x1);
    dx = [x2 x3 ones(size(x3(:,1)))];

    xlswrite(xfile,dx');
    xlswrite(yfile,y');

    x1=dx';
    y1=y';

    for k=1:n_e
    y_new = neural1 (x1,y1,nmatrix(k));
    xlswrite(outfile,y_new',2*k-1);
    [b,bint,r,rint,stats] = regress (y_new',dx); % regression parameters calculated for
all values
```

```
    [r1,m1,b1] = regression (y1,y_new,'one');
    q = [nmatrix(k) r1];
    xlswrite(outfile,q,2*k);
    if k==ceil(n_e/2) % full equation computed for middle value
    %o = xloutput (outfile0,b,stats(1),c,nmatrix(k));
    m = b;
    r = stats(1);
    n = nmatrix(k);
    %o = xloutput (c);

a1 = m(1);
a2 = m(2:c(2));
a3 = m(c(2)+1:c(3));
a4 = m(c(3)+1:c(4));
a5 = m(c(4)+1:c(5));
a6 = m(c(6));
%a6 = m(c(5)+1:c(6));
%a7 = m(c(7));
xlswrite(outfile0,a1,'Age');
xlswrite(outfile0,a2,'Season');
xlswrite(outfile0,a3,'Time');
xlswrite(outfile0,a4,'Reason');
xlswrite(outfile0,a5,'Affected Organ');
%xlswrite(file,a6,'Organ');
xlswrite(outfile0,a6,'Constant');
xlswrite(outfile0,r,'R^2');
xlswrite(outfile0,n,'number of layers');
    end
    end
    pfm = 0;
```

8.2 neural1.m

```
function y = neural1 (x,t,n)
setdemorandstream(689271451)
net = fitnet (n);
[net,tr] = train(net,x,t);
y = net (x);
```

8.3 xlinput.m (same as 6.2)

8.4 xloutput (same as 6.3

APPENDIX D

CODES FOR FUZZY LOGIC

9.1 main file (master.m)

```
function pfm = master (inputfile,rulefile,outputfile)
    % x contains independent parameters' columns
    % y is the response column
    [x,y] = xlinput (inputfile);
    % matrix obtained
    % M1 - M4: matrices for average values of y corresponding to parameters
    % R: matrix for rules
     M1 = zeros (4,2);
     M2 = zeros (3,2);
     M3 = zeros (14,2);
     M4 = zeros (8,2);
    %M1 - 4 significant seasons 1,2,3,4
    j1 = 0;
    for j=1:4
    x1 = find (x(:,1)==j);
    x0 = size(x1,1);
    if (x0>29) %minimum 30 elements
        m = mean(y(x1));
    M1(j-j1,:) = [m,j];
    else
    M1(j-j1,:) = [];
    j1=j1+1;
    end
```

```
j1 = 0;
%M2 - 3 time shifts 1-3
for j=1:3
x1 = find (x(:,2)==j);
x0 = size(x1,1);
if (x0>29) %minimum 30 elements
m = mean(y(x1));
M2(j-j1,:) = [m,j];
else
M2(j-j1,:) = [];
j1=j1+1;
end
j1 = 0;
%M3 - 14 reasons
for j=1:14
x1 = find (x(:,3)==j);
x0 = size(x1,1);
if (x0>29) %minimum 30 elements
m = mean(y(x1));
M3(j-j1,:) = [m,j];
else
M3(j-j1,:) = [];
j1=j1+1;
end
j1 = 0;
%M4 - 8 organs
for j=1:8
x1 = find (x(:,4)==j);
x0 = size(x1,1);
if (x0>29) %minimum 30 elements
m = mean(y(x1));
```

```
M4(j-j1,:) = [m,j];
else
M4(j-j1,:) = [];
j1=j1+1;
end
M5 = sumfour (M1,M2,M3,M4); %computes sum of numbers
val = M5(:,5); % value column
nv = numel(val);
val1 = sortrows(val);
ne = ceil(nv/3); %no. of elements
lo = val1(ne);
hi = val1(nv-ne);
dec = zeros(nv);
for i=1:nv
   if(val(i)<lo)
      dec(i) = 1; %low workday loss
   else
      if(val(i)>hi)
         dec(i) = 3;
      else
         dec(i) = 2;
      end
M5 = [M5 dec]; %decision appended to M5
xlswrite(rulefile,M5,'Fuzzy rules');
xlswrite(outputfile,M1,'1');
xlswrite(outputfile,M2,'2');
xlswrite(outputfile,M3,'3');
xlswrite(outputfile,M4,'4');
pfm = 0;
```

9.2 sumfour.m

```
function z = sumfour (a,b,c,d)
sa = size(a,1);
sb = size(b,1);
sc = size(c,1);
sd = size(d,1);
z = zeros(sa*sb*sc*sd,5);
s_r = 1; %row counter
for p=1:sa
    for q=1:sb
        for r=1:sc
            for s=1:sd
                z(s_r,1) = a(p,2);
                z(s_r,2) = b(q,2);
                z(s_r,3) = c(r,2);
                z(s_r,4) = d(s,2);
                z(s_r,5) = a(p,1)+b(q,1)+c(r,1)+d(s,1);
                s_r=s_r+1;

            end
```

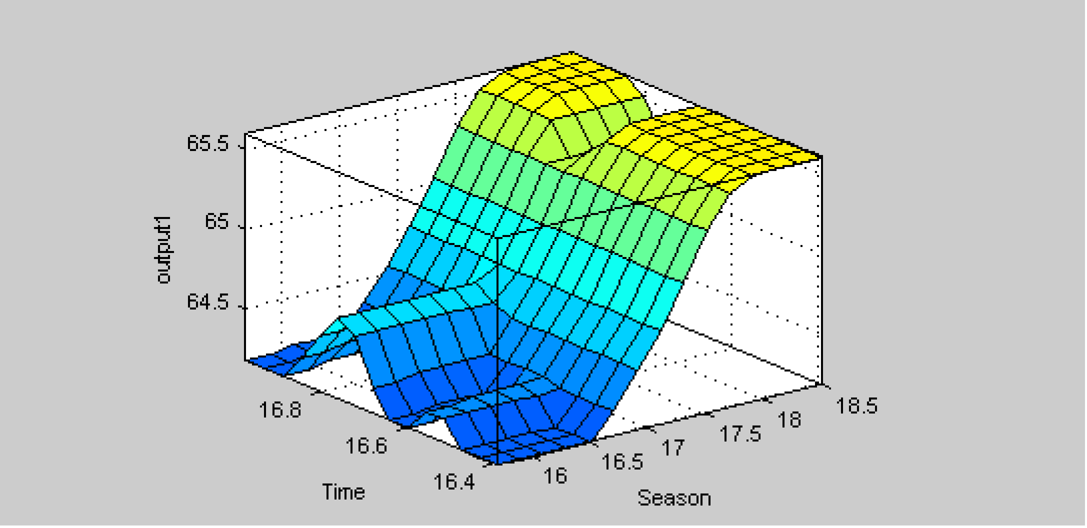
9.3 xlinput.m (same as 6.2)

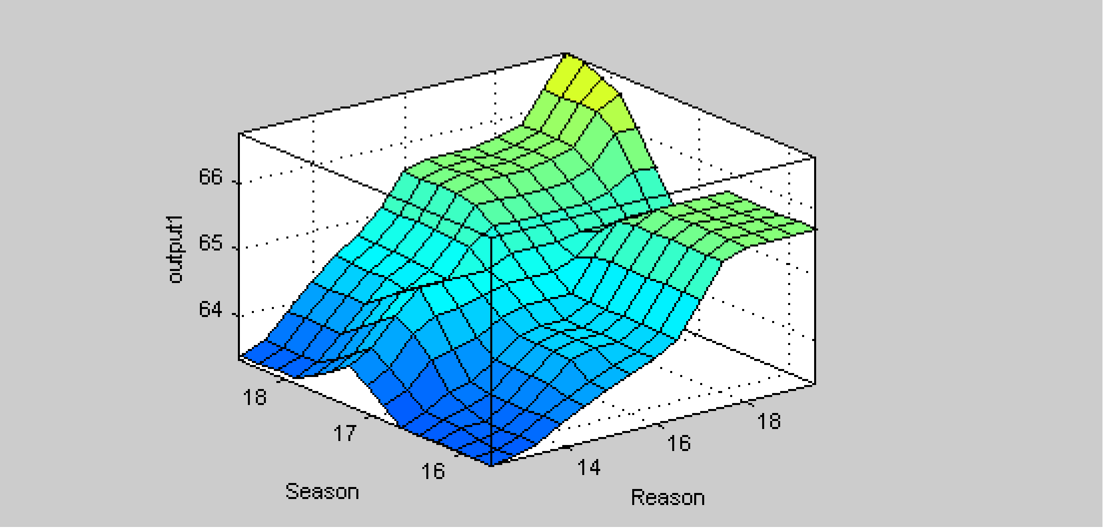
9.4 xloutput (same as 6.3)

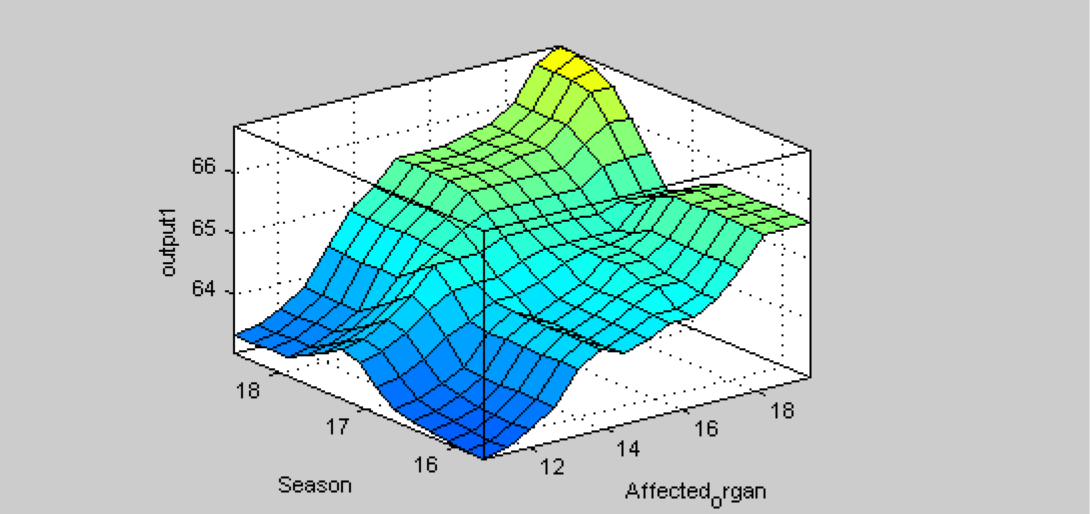
APPENDIX E

SURFACE GRAPHICS

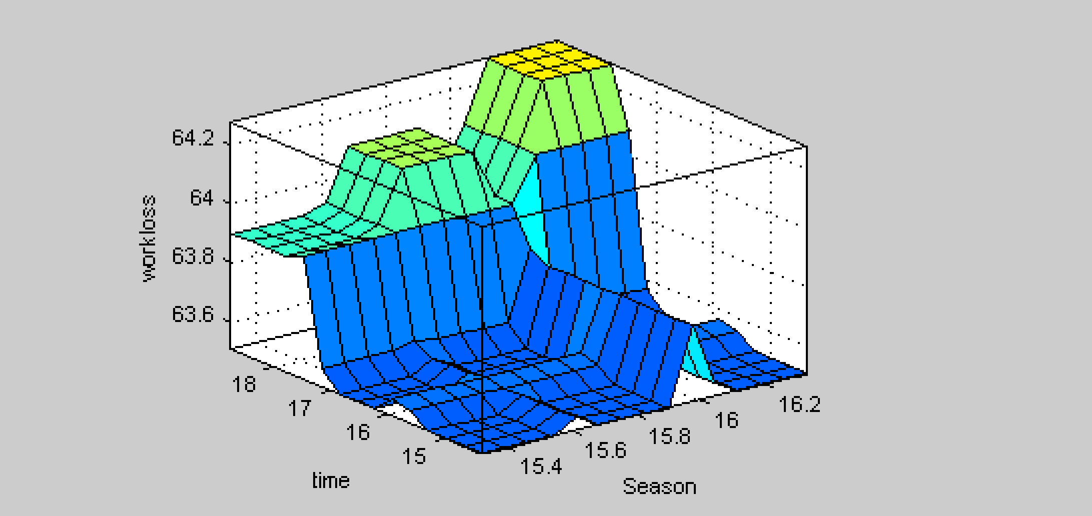
1. Surface graphics for ELI mines

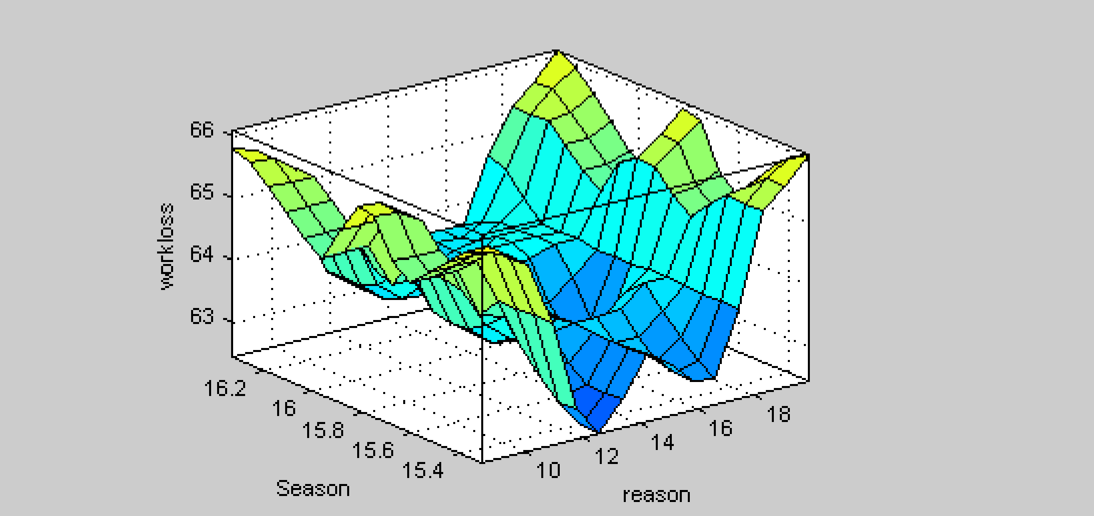


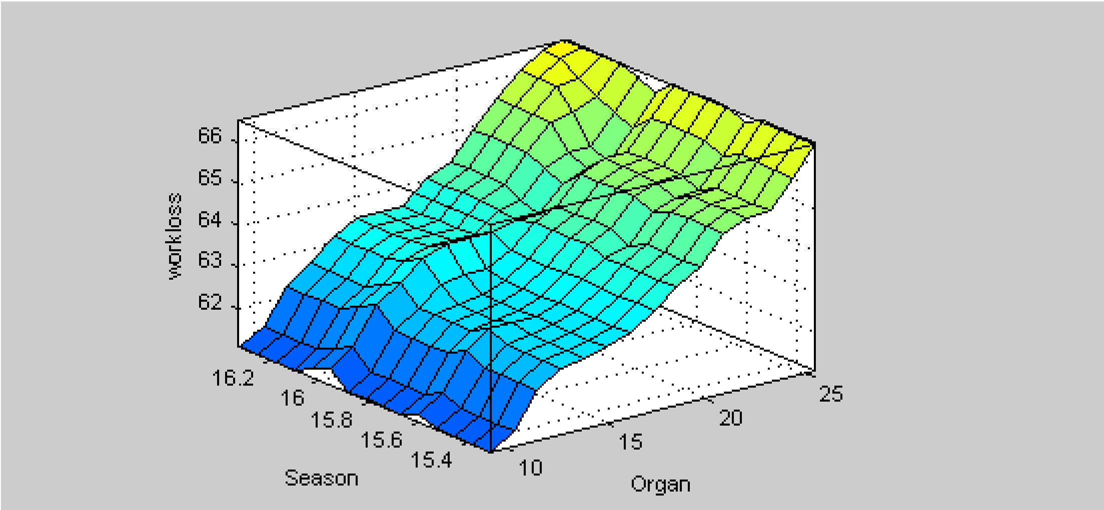




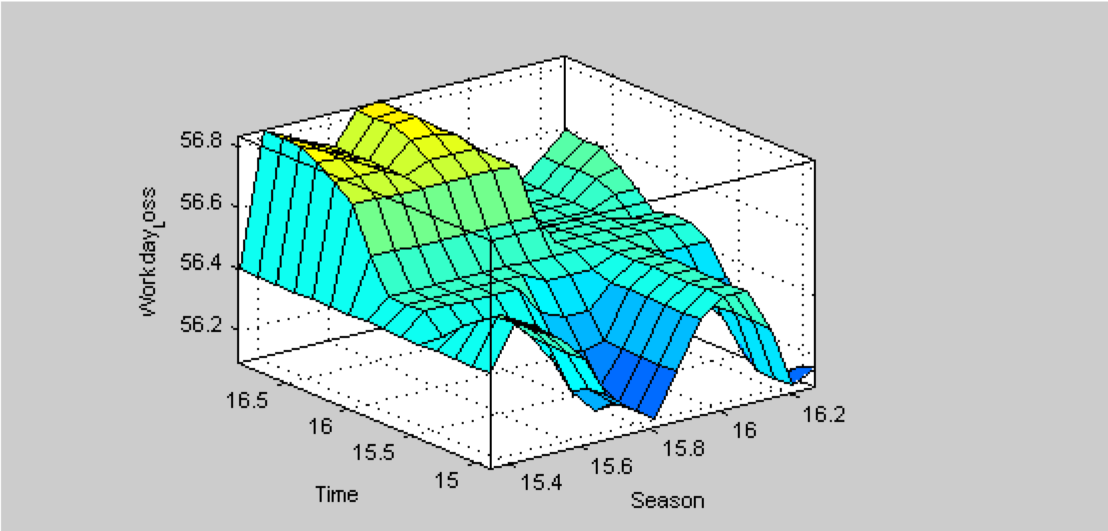
2. Surface graphics for GLI mines

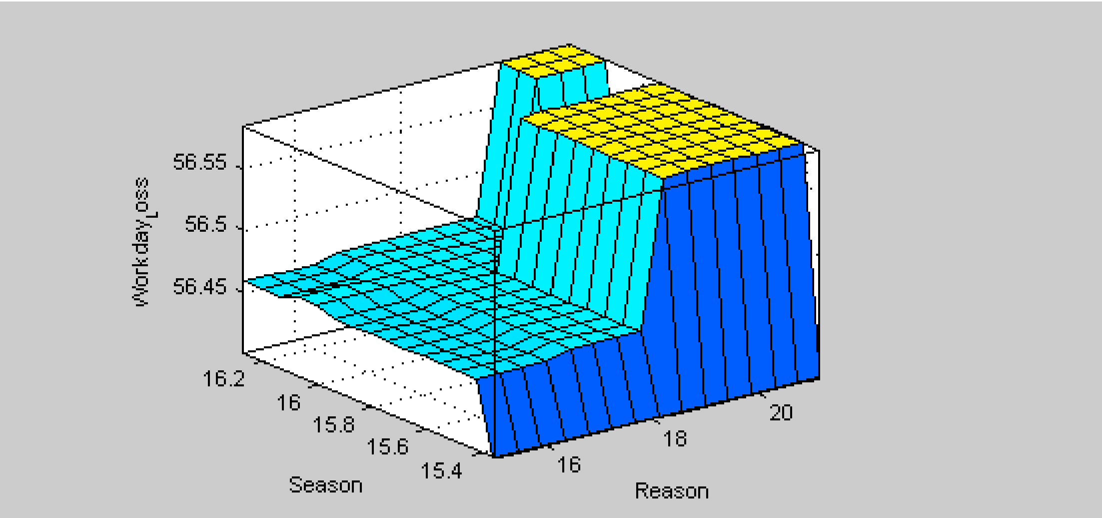


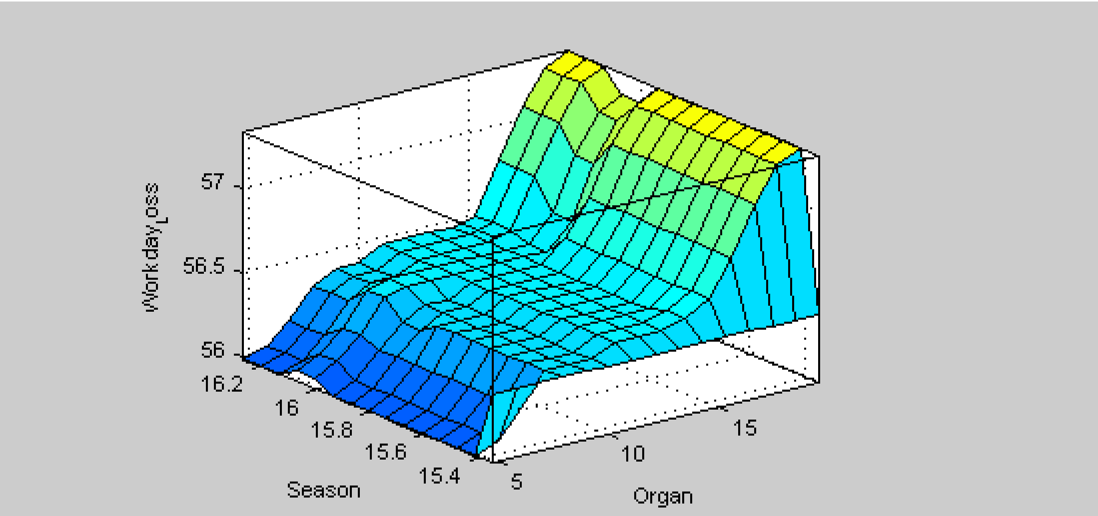




3. Surface graphics for TTK mines




Workday_Loss
56.55
56.5
56.45
16.2
16
15.8
15.6
15.4
Season
16
18
20
Reason


Workday_Loss
57
56.5
56
16.2
16
15.8
15.6
15.4
Season
5
10
15
Organ

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Eratak, Özlem Deniz

Nationality: Turkish (TC)

Date and Place of Birth: 16 January 1978, Ankara

Marital Status: Single

Phone: +90 505 419 52 50

email: deniz.eratak@gmail.com

EDUCATION

Degree	**Institution**	**Year of Graduation**
MS METU	Mining Engineering	2005
BS ITU	Mining Engineering	1999
High School	Arı Fen High School	1995

WORK EXPERIENCE

Year	**Place**	**Enrollment**
2011-Present	DuPont Performance Coatings	Regional EHS Manager
2005 – 2011	Ministry of Labour	OHS Expert
2002 – 2005	Megapol Construction	Project Manager
2000 – 2002	Park Holding	Mining Engineer

FOREIGN LANGUAGES

Advanced English

PUBLICATIONS

1. Ö.D.Eratak, İ Acar, F.Başayar (2011) Effects of Work Related Psychosocial Factors on Coal Miners, , XIX World Congress on Safety and Health.
2. , İ.Çelik, Ö.D.Eratak and A.R. Ergun (2010) Risk Assessment Approach in Underground Coal Mining, Health and Safety Congress
3. Ç.Hoşten, Ö.D.Eratak (2006) A novel simple method for contact angle determination of particulate solids, XXIII International Mineral Processing Congress.
4. G.Atesok, S.Celik, F.Boylu, Ö.D.Eratak (2000) Carrier Flotation for Desulfurization and Deashing of Difficult-to-float Coals, , Fuel and Energy.

HOBBIES

Music, Movies, Photography, Computer Games