INTERNATIONAL ISLAMIC UNIVERSITY ISLAMABAD

Unraveling the Underlying Factors of Cognitive Failures in Construction Workers: A Safety-Centric Exploration

THESIS SUBMITTED TO

DEPARTMENT OF CIVIL ENGINEERING INTERNATIONAL ISLAMIC UNIVERSITY, ISLAMABAD

FOR THE AWARD OF DEGREE OF

MASTER OF SCIENCE

IN

CIVIL ENGINEERING

Submitted By

MUHAMMAD ASGHAR **15-FET/MSCE/F22**

December 2025

Dedication

I dedicate this thesis with profound gratitude to my beloved parents, whose unwavering support, selfless sacrifices, and constant prayers have been the bedrock of my journey. They have always prioritized my future, making countless sacrifices for my well-being and instilling in me strong moral values. I also extend this dedication to my esteemed teachers, whose invaluable guidance and unwavering encouragement have been a constant source of inspiration and knowledge throughout my academic pursuit.

FINAL APPROVAL

It is certificate that we have read the thesis submitted by Muhammad Asghar and it is our judgment that this thesis is of sufficient standard to warrant its acceptance by the International Islamic University, Islamabad for the MS Degree in Civil Engineering

COMMITTEE

External Examiner

Rashid Ullah

Internal Examiner

Supervisor

Chairman

Author's Declaration

I, **Muhammad Asghar** hereby state that my MS thesis titled "Unraveling the Underlying Factors of Cognitive Failures in Construction Workers: A Safety-Centric Exploration" is my work and has not been submitted previously by me for taking any degree from International Islamic University, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

Muhammad Asghar
15-FET/MSCE/F22

Plagiarism Undertaking

I solemnly declare that the research work presented in this thesis titled "Unraveling the Underlying Factors of Cognitive Failures in Construction Workers: A Safety-Centric Exploration" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and International Islamic University, Islamabad towards plagiarism. Therefore, I as an author of the above-titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above-titled thesis even after awarding of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

____

Muhammad Asghar
15-FET/MSCE/F22

Acknowledgments

I begin by expressing my profound gratitude to **ALLAH ALMIGHTY**, whose divine grace and guidance provided me with the strength, wisdom, and perseverance necessary to complete this research endeavor. I am deeply indebted to my beloved parents for their unwavering support, constant prayers, and endless encouragement, which have consistently formed the foundation of my achievements. Their constant prayers for my success have been a tremendous source of motivation.

I extend my sincere appreciation to **Engr. Muhammad Abu Bakar Tariq** as my supervisor and **Dr. Arsalan Khan** for their expert guidance, insightful remarks, and invaluable technical support throughout this research. His unwavering enthusiasm and keen interest in my progress were instrumental in keeping me engaged and motivated, transforming this challenging process into a fulfilling experience. It has been a privilege to have worked under his mentorship.

Table of Contents

Table of Contents....2

List of Figures....4

List of Tables....4

Abstract....1

Chapter 1 : Introduction....3

1.1. Research Problem....5

1.2. Objectives of the Research....6

1.3. Significance of the Research....7

1.3.1 Enhancing Construction Safety....7

1.3.2 Improved Worker Well-being....7

1.3.3 Improved Project Efficiency and Cost Management....8

1.3.4 Closing Research Gaps....8

1.3.5 Practical Applications....8

1.3.6 Anticipating Future Challenges....8

Chapter 2 : Literature Review....9

2.1. Cognitive Failures and Safety Risks in Construction....9

2.2. Safety Cognition and Risk Awareness in Construction....9

2.3. Situation Awareness and Cognitive Load in Construction Safety....10

2.4. Understanding Risky Behaviors in Construction Workers....10

2.5. Causation Patterns and Behavioral Influences in Construction....10

2.6. Group Safety Management Model for Construction Workers....11

2.7. Impact of Stress on Unsafe Behavior in Construction Workers 12
2.8. Novelty of work 12
Chapter 3 : Research Methodology 16
3.1. Literature Review and Factor Identification 18
3.2. Development of Priority Matrix...18
3.3. Questionnaire Development..19
3.4. Data Collection 20
3.5. Data Analysis...21
3.5.1. Statistical Techniques 23
3.5.2. Machine Learning Predictive Techniques 27
3.5.3. Performance Evaluation Metrics 31
Chapter 4 : Analysis and Results...34
4.1 Descriptive Analysis 34
4.2 Correlation Analysis 39
4.3 Statistical Models..41
4.3.1 Linear Regression 41
4.3.2 Binary Logistic Regression 44
4.4 Machine Learning Predictive Models...47
4.4.1 Support Vector Machines (SVM)...48
4.4.2 Random Forest (RF) 50
4.4.3 Gradient Boosting...51
4.4.4 Optimal Performing ML Model Based on Data 53
4.4.5 Feature Importance 54

Chapter 5 : Discussion 56

Chapter 6 : Conclusions & Recommendations 61

References 63

List of Figures

Figure 1: Methodology of the study 17

Figure 2:Selection of Cognitive Factors 19

Figure 3:Age Distribution of Respondents 35

Figure 4: Distribution of Respondents on Basis of Work Experience 36

Figure 5: Distribution of respondents on basis of Education Level 38

Figure 6: Correlation Heat Map 41

Figure 7: Feature Importance ranking based on AUC decrease 56

List of Tables

Table 1: Research Papers and Cognitive Factors 13

Table 2: Performance Evaluation Parameters 31

Table 3: Age Distribution of Respondents 34

Table 4: Distribution of Respondents on Basis of Work Experience 36

Table 5: Distribution of respondents on basis of Education Level 37

Table 6: Descriptive analysis of all variables 38

Table 7: Correlation Matrix 40

Table 8: Linear Regression Analysis 42

Table 9: Binary Logistic Regression Analysis 44

Table 10: Results of SVM Analysis 49

Table 11: Confusion Matrix for SVM Model 49

Table 12: Results of Random Forest Analysis 50

Table 13: Confusion Matrix for Random Forest Model 51

Table 14: Results of Gradient Boosting Analysis 52

Table 15: Confusion Matrix for Gradient Boosting Model 53

Abstract

The construction industry exhibits major safety concerns since cognitive mistakes among employees lead to a great frequency of accidents and dangerous actions. This study investigates the main psychological, behavioral, and environmental elements causing cognitive failure in the construction sector. To find pertinent factors, a multi-stage method was used starting with literature study to identify 30 factors of cognitive failures, followed by expert opinion of ten seasoned safety experts who used an organized priority matrix to assist ranking top 10 the most important cognitive elements. Drawing on the results of the aforementioned steps, a questionnaire survey was drafted. The survey was conducted among the construction employees of different construction sites and 500 valid responses were collected from those working in basic labor to engineering. Both traditional statistical tools and contemporary machine learning techniques were used in analysis of the responses. Linear regression first enabled investigation of how various factors affect the degree of cognitive deficits. The likelihood of these failures was thereafter categorized using binary logistic regression. Predictive accuracy was raised by advanced techniques including Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting (GB); of these, Gradient Boosting fared the best in terms of both prediction and finding the most essential variables. While fatigue, education, and employment history were most obvious forecasts, stress, safety knowledge, and awareness of safety were found to be the most important contributions. These findings underline the need for focused safety training, stress-reducing initiatives, and continuous observation to lower cognitive mistakes on construction sites. All things taken into account, the study provides important guidance for safety managers, lawmakers, and construction businesses by combining professional viewpoints with

modern analytical methodologies, therefore offering a data-driven foundation for improving occupational safety.

Keywords: Safety During Construction, Cognitive Impairments, Unsafe Conduct, SVM, Support Vector Machine Forecasting, Awareness of Safety, Knowledge About Safety, Levels of Stress, Reduction of Risk, Workplace Security, Preventing Accidents, Data-Driven Methodology.

.

Chapter 1 : Introduction

In order to guarantee workers' safety and welfare in construction industry, the cognitive aspects influencing their behaviour need to be given great thought. Though crucial, not much is understood about how these elements influence safety, especially in developing countries like Pakistan (Deng et al., 2022). This study seeks to address that gap by looking at the manners in which behavioral, psychological in nature and external variables lead to thinking breakdowns among construction workers. Usually showing up as concentration, memory, and decision-making mistakes, these disruptions increase the likelihood of unsafe behaviour and workplace mishaps.

Construction industry is regularly placed among the riskiest sectors given the stated annual death and injury count (Deng et al., 2022; Ju et al., 2013). Studies indicate that between 80% and 90% of these events are brought on by employees' own risky behavior (Razavi, 2019;Liu et al., 2023). In this setting, cognitive failure describes typical mistakes that hinder the detection of hazards, the processing of data, or the execution of safety procedures. Among the usual reasons of these errors are stress, fatigue, lack of experience, or ignorance of safety precautions. Understanding these cognitive challenges can help one to design more targeted and successful safety solutions (Razavi, 2019;Liu et al., 2023).

Previous research has pointed up many factors influencing cognitive performance in construction industry. These comprise human elements including stress, tiredness, education, and work experience as well as outside pressures including harsh weather and limited project times (Kim et al., 2022). Furthermore, greatly influencing employee behaviour on the job are safety training, awareness initiatives, and job culture. Development of more focused and effective safety-related strategies depends on an understanding of and identification of these key elements.

This study uses a pragmatic, evidence-based, data-driven approach. First a thorough review of the research was done to identify common cognitive failure characteristics in construction industry. A group of ten experienced safety professionals was consulted to assist in developing a priority matrix so that the findings would be localized. This guaranteed that the found factors were relevant and greatly influenced Pakistan's construction sector. With the help of this professional advice, a focused list of 10 significant cognitive aspects for further investigation was developed.

Designed with the authorized set of ten criteria, a systematic survey was developed and distributed among several construction projects including highways, dam sites, bridge construction, and residential projects. Professionals at several levels—from engineers to laborers, foremen to site supervisors—provided direct responses. Notwithstanding early participant reluctance and administrative difficulties, the data collecting technique yielded a strong 500 valid response count for analysis.

The data was examined using an analytical method comprising of multiple steps. The method started with linear regression to look at how factors including stress, fatigue, and safety consciousness influenced the level of cognitive failures.

Binary logistic regression was applied to identify key predictors significantly associated with the likelihood of cognitive failure among construction workers. Furthermore, it was used to increase forecast accuracy and evaluate the relative relevance of every variable were machine learning methods including Support Vector Machine (SVM), Random Forest (RF), and Gradient

Boosting (GB), so improving general safety results in the construction sector. Occupational mishaps also had a significant impact on worker safety, project timelines, and overall productivity.

This methodical approach offers major developments to construction safety theory and practice by means of the integration of expert knowledge, traditional statistical models, and advanced machine learning techniques. The final objective of the study was to create a structure that would help legislators, safety managers, and construction businesses in identifying key factors contributing to cognitive failure and using them to inform targeted safety interventions. This helps the research support programs aimed at reducing on-site accidents and enhancing general safety standards in the construction industry.

1.1. Research Problem

1. Particularly in China, there have been a lot more accidents and deaths in the construction sector recently. For example; in 2015 there were 442 accidents and 554 deaths; in 2019 there were 773 accidents and 904 deaths in China. Globally, too, trends show the ongoing hazards in construction. With around 1,300 deaths yearly in the European Union, construction industry is the most dangerous industry (U.S. Bureau of Labor Statistics, 2021). In the United Kingdom, 10% of significant injuries and 31% of occupational fatalities come from construction (Health and Safety Executive (HSE) UK, 2024). Comparably, in Australia the construction sector recorded a 6.2 per 100,000 worker mortality rate in 2018 (Safe Work Australia, 2018) These concerning numbers emphasize how urgently safety issues in the construction industry ought to be addressed. Research studies repeatedly show that the main reason for these deaths is the risky behaviour of construction professionals. Based on an

analysis of 75,000 events, Heinrich's accident causation theory—which predicated on human causes—saw 88% of accidents having human elements (Razavi, 2019).

In Pakistan, the construction industry has also been repeatedly identified as one of the highest-risk sectors. Reports from the Pakistan Labour Force Survey and PEC highlight frequent accidents involving falls, electrocution, equipment failures, and structural collapses. These recurring incidents demonstrate the urgent need to understand cognitive failures and safety behaviour within Pakistan's construction environment.

2. Similarly, studies conducted in Finland found that risky human behaviour accounted for 84–94% of work-related deaths. Researchers all around have studied construction accidents in great detail, and they have repeatedly found that the main reason is risky behavior (Kines et al., 2010; Pui et al., 2000). Moreover, data from other nations, such the USA, UK, and China, show seriousness of the issue since the construction sector is clearly connected to a large number of occupational deaths (Dongping et al., 2016).
3. Not exclusive to any one country, the worldwide issue of rising numbers of accidents involving construction is one of development. Dealing with safety concerns become more important given the great risk associated with construction activities and the often-shifting site conditions (Zhang et al., 2019). As seen in (Liu et al., 2023), this study emphasizes the need of robust safety control policies to reduce the effect of workers' dangerous behaviour. Raising security criteria in the construction industry depends on a strong awareness of the fundamental causes of these dangerous activities and their direct influence on general safety performance.

1.2. Objectives of the Research

The study aims at the following goals:

1. To methodically pinpoint and rank the main cognitive elements influencing worker safety in the construction industry.
2. To quantitatively assess how the identified cognitive factors, affect the cognitive performance of construction workers.
3. To establish cognitive failure factors based on a data-driven approach to improve health and safety in the construction industry.

1.3. Significance of the Research

In the context of construction safety and the larger subject of occupational health and safety, the suggested study examining the variables influencing the cognitive abilities of construction workers is highly significant. Numerous noteworthy benefits and justifications are expected by tackling this study problem.

1.3.1 Enhancing Construction Safety

The construction sector is well known for having a high accident and injury rate. The progress of construction safety greatly benefits from this research. Enhanced comprehension of the elements impacting employees' cognitive abilities can result in safer safety practices, which in turn can lower the number of mishaps and deaths (Deng et al., 2023) (Ding et al., 2018) (S. Guo et al., 2018).

1.3.2 Improved Worker Well-being

The cognitive health of construction workers has a direct effect on their wellbeing. In addition to saving lives, a safer workplace with fewer accidents improves employees' general physical and mental well-being (Tong et al., 2023). It is anticipated that the research findings will result in initiatives that can improve construction workers’ well-being (Deng et al., 2023).

1.3.3 Improved Project Efficiency and Cost Management

Costs and schedules for projects might be affected by safety issues. Construction processes become more efficient in a safer work environment, which lowers delays and cost overruns (Liang et al., 2022). It is expected that this research will improve construction enterprises' profitability and competitiveness.

1.3.4 Closing Research Gaps

Although many aspects of construction safety—such as site hazards, unsafe behaviours, PPE use, and environmental risks—have been studied, the role of cognitive factors (stress, fatigue, attention, safety awareness, experience, and knowledge) is still not fully explored.

1.3.5 Practical Applications

It is anticipated that the research findings will have immediate use in the construction sector. The research findings can be used by safety managers, legislators, and construction firms to create and carry out focused safety initiatives, training courses, and interventions.

1.3.6 Anticipating Future Challenges

This research can assist the construction industry in anticipating future safety concerns and proactively addressing them, as the industry is always changing. It would shed light on how to modify safety procedures to account for evolving work environments.

In summary, the importance of this research lies in its ability to close important knowledge gaps, boost industry competitiveness, and improve the safety and well-being of construction workers. The findings have the potential to impact real-world implementations and foster a proactive, safer, and more effective construction industry.

Chapter 2 : Literature Review

2.1. Cognitive Failures and Safety Risks in Construction

As construction workers are involved in safety mishaps, the construction industry is considered dangerous on a global scale. The current literature on unsafe behavior does not provide a thorough justification. This study presents a cognitive failure model in response. It lists the key cognitive factors causing cognitive failure thus resulting in unsafe behavior. Future studies ought to gather more data and examine more cognitive aspects (Deng et al., 2022). Because construction has higher injury rates than other industries, it is imperative to improve safety (Molen et al., 2005). Destructive harm persists in spite of multiple safety research and policy suggestions (Razavi, 2019).

2.2. Safety Cognition and Risk Awareness in Construction

Construction workers' unsafe behavior, a global challenge tied to human cognition, prompts the need for focused research. The study conducted by (Liu et al., 2023) summarizes and reviews antecedents of construction workers' safety cognition, proposing a four-level model and exploring relationships among these factors. Suggestions for future research include a holistic perspective, causal antecedents, and mechanisms of cognitive failure, ultimately enhancing our understanding of safety cognition and guiding safety management.

2.3. Situation Awareness and Cognitive Load in Construction Safety

Construction is highly risky due to unsafe worker behavior tied to their ability to recognize and respond to hazards (Silva et al., 2004; Edelson et al., 2009; Askaripoor & Jafari, 2015). Prior research explored workers' perception and comprehension of safety information, particularly in relation to their workload and situation awareness (SA). In real construction settings, workers' situation awareness was assessed, with a focus on measuring perception (Level-measuring perception (Level-1 SA: noticing and identifying safety-related information) and comprehension (Level-2 SA: understanding what that information means for the task and potential risks) as they carried out construction tasks. The results showed significant differences in safety information perception and comprehension, with mental load negatively impacting SA, validating the impact of workload on workers' perception and comprehension at actual construction sites, which can enhance safety management guideline (Kim et al., 2022).

2.4. Understanding Risky Behaviors in Construction Workers

Because construction accidents result in a significant mortality toll, it is necessary to look into the risky practices of construction workers (Shakerian et al., 2021). Eleven characteristics were used to create a questionnaire based on a cognitive model in order to better understand these behaviors. The questionnaire proved to have strong validity and reliability after verification, which makes it useful for determining the reasons behind risky behaviors in construction workers(Deng et al., 2021)

2.5. Causation Patterns and Behavioral Influences in Construction

Patterns of causation can be found by analyzing close calls with falls (Weili et al., 2018). The main causes of unsafe behavior include habits, motivation, subjective norms, and perceived behavioral control. This offers a novel viewpoint on construction safety and aids managers in assessing the aggregate effect of causative elements on employees' safety conduct (Mohajeri et al., 2022).

2.6. Group Safety Management Model for Construction Workers

The study conducted by (Deng et al., 2023) took individual features into account, a group safety management model for construction workers was developed. Ten-factor cognitive model was created, verifying a questionnaire and classifying workers into four groups. Employees who belong to the cognitive failure type have low cognitive capacity, whereas those who belong to the cognitive excellent type have superior cognitive ability and less of a tendency for risky behaviors. Employees with no fear type have poor cognitive ability; they may struggle to find knowledge and select coping mechanisms, and employees with knowing offender type may struggle to select coping mechanisms. Because of the small data coverage and scope, additional research and empirical validation are required. However, this study aims to provide focused management measures for each category and provides theoretical evidence for personalized unsafe behavior management.

2.7. Impact of Psychological Factors on Construction Workers' Safety Performance During COVID

With an emphasis on COVID-19, the effects of psychological elements on construction workers' performance in terms of safety were investigated. Among these characteristics were interpersonal conflict, role ambiguity, autonomy, social support, workplace stress, and work-family conflict. The findings indicate that while autonomy and social support have positive benefits on safety performance, work-family conflict, role ambiguity, interpersonal conflict, and job stress all have negative consequences." To improve workers' involvement and well-being in the sector, it was emphasized that these psychological variables should be taken into consideration while establishing regulations to regulate construction safety(Tong et al., 2023)

2.8. Impact of Stress on Unsafe Behavior in Construction Workers

Studying the impact of stress on the risky behavior of construction workers revealed that stress can both directly cause risky behavior and have an indirect effect on safety through cognitive processes (Austin et al., 1996; Shakerian et al., 2021). It was possible to gain an understanding of the relationships between stress, cognition, and safety as well as useful suggestions for safety management in the construction sector (Liang et al., 2022).

2.9. Novelty of work

This study provides a helpful, evidence-based paradigm to help one better understand and lower cognitive failures among construction workers (Bohm & Harris, 2010; Choudhry & Fang, 2008). This issue has not gained much attention in underdeveloped countries like Pakistan. Unlike depending just on mathematical models, the study ranks the most essential cognitive aspects affecting unsafe behavior using an organized priority matrix and combines expert opinions from experienced safety professionals. Further distinguishes the study is its coverage of a wide range of real-world construction sites—including buildings, bridges, roads, and dams—as well as its broadened sample of 500 respondents, which consists of engineers, supervisors, and workers. Important cognitive components were turned into measurable variables using a well-crafted questionnaire, therefore guaranteeing accuracy and applicability. The collected data was analyzed using a mix of modern machine learning algorithms like the Support Vector Machine, Random Forest, and Gradient Boosting together with conventional statistical methods; the latter displayed the best prediction accuracy. Apart from raising the validity of the results, this combined analytical approach provides helpful guidance for legislators and safety professionals. Ultimately, the study reveals a vital link between cognitive neuroscience and on-site oversight of safety, therefore offering a new path for enhancing safety outcomes in the construction industry.

Table 1: Research Papers and Cognitive Factors

Sr. No	Research Paper	Years of Publication	Cognitive factors
1	A Cognitive Failure Model of Construction Workers Unsafe Behavior (Deng et al., 2022)	2022	Safety Vigilance, Hazard Identification, Safety Knowledge, Safety Behavior Attitude, Professional Skills
2	Antecedents of Construction Workers Safety Cognition (Liu et al., 2023)	2023	Age, Education Level, Personality, Job Position, Marital Status, Smoking, Sleep, Fatigue Level, Experience, Time Pressure, Weather Conditions, Reward/Penalty for Safe/Unsafe Behaviors and Performances
3	A Review of Social, Physiological, and Cognitive Factors Affecting Construction Safety (Razavi, 2019)	2019	Loudness at Site Conditions, Lack of Attention
4	Construction Workers	2022	High Load Physical

	Awareness of Safety Information Depending on Physical and Mental Load (Kim et al., 2022)		Exertion
5	Development and Validation of a Cognitive Model-Based Novel Questionnaire for Measuring Potential Unsafe Behaviors of Construction Workers (Deng et al., 2021)	2021	Safety Awareness
6	Discovering Causality Patterns of Unsafe Behavior Leading to Fall Hazards (Mohajeri et al., 2022)	2022	Habits, Motivation, Perceived Behavioral Control
7	Group Management Model for Construction Workers Unsafe Behavior Based on Cognitive Process Model (Deng et al., 2023)	2023	Work Skills
8	Psychosocial Factors for Safety Performance of Construction Workers: Taking Stock and Looking Forward (Tong et al., 2023)	2023	Work Schedule (Shift work, night shifts, and long or unsociable hours), Interpersonal Relationship (Social or physical isolation, interpersonal conflict)
9	Unveiling the Mechanism of	2022	Stress

	Construction Workers Unsafe Behavior from an Occupational Stress Perspective (Liang et al., 2022)		
10	How do psychological cognition and institutional environment affect the unsafe behavior (Yuan et al., 2022)	2022	Safety Trainings, Safety Beliefs

Chapter 3 : Research Methodology

The methodical and orderly approach applied in the study is shown in the accompanying flowchart (Figure 1). Starting with literature review to find factors contributing to cognitive failure of construction workers, followed by a priority matrix to rank top ten primary factors of cognitive failure by consulting with experts. Afterwards, based on the expert opinion having top ten cognitive failure factors a questionnaire was developed, survey was conducted in different construction sites, and 500 valid worker responses were gathered. After data preparation and cleaning, the analysis was done in phases, starting with descriptive statistics, then using traditional models including logistic regression and regression models, and lastly the integration of Machine Learning (ML) algorithms including Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting (GB). Model results were compared to assessing performance of each ML model and pinpoint the most important predictors for cognitive failure, which at last produced practical, evidence-based safety recommendations.

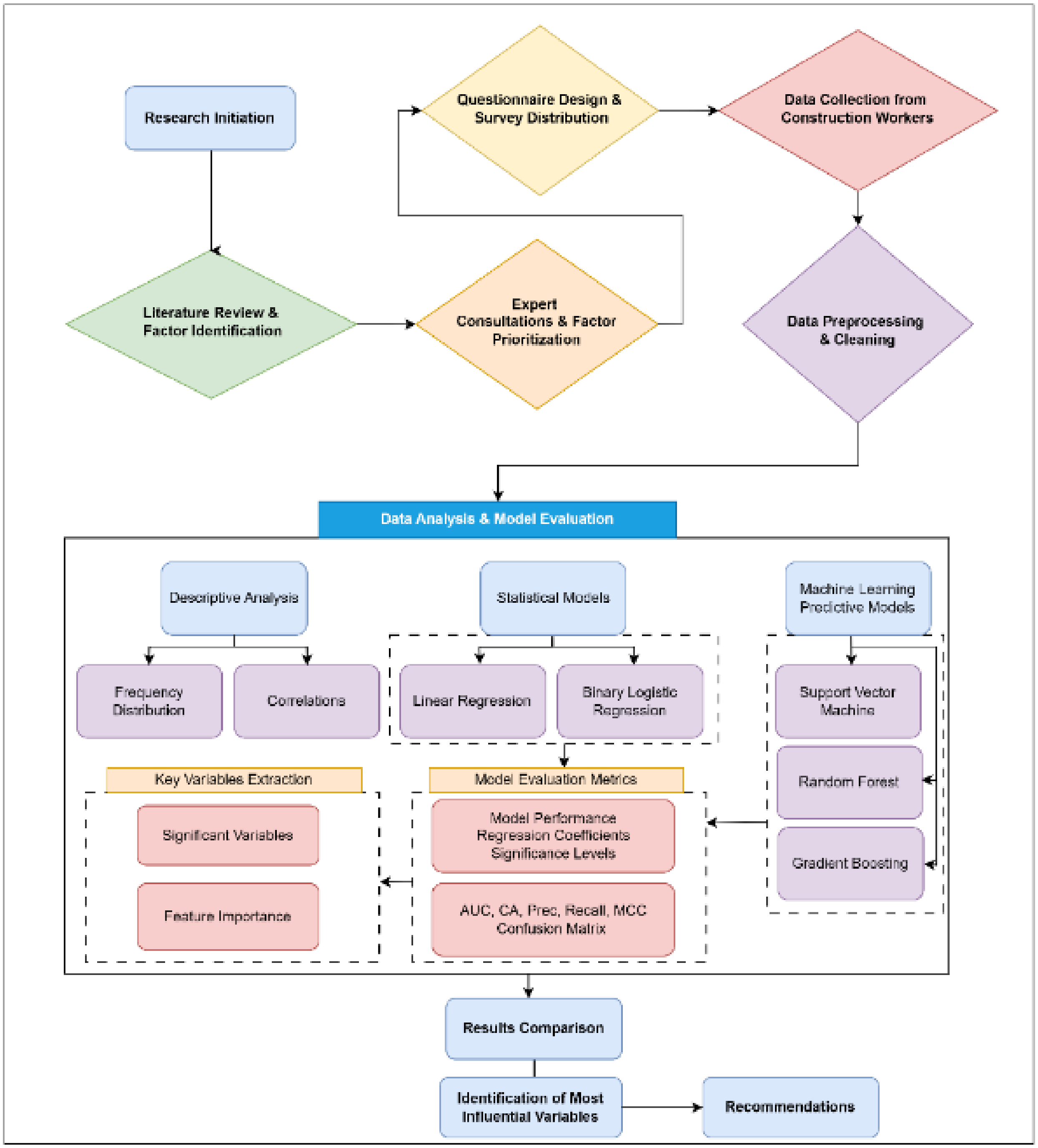


Figure 1: Methodology of the study

3.1. Literature Review and Factor Identification

The study started with a detailed literature review to determine factors that construction workers experience when they experience cognitive failures. This initial phase of the research consisted of a comprehensive review of the existing literature leads to identification of 30 cognitive failure factors, to provide a basis of the research area before moving to the next stages.

3.2. Development of Priority Matrix

Ten experienced safety managers consulted in developing a Priority Matrix, building upon insights derived from the literature review. These experts identified and prioritize the ten most critical factors specific to Pakistan’s construction sector. To ensure an efficient and inclusive selection process, Google Forms was utilized for data collection and evaluation. Ten experienced safety managers, each having 5–10 years of construction experience, were given all 30 factors identified from the literature review and were asked to rank them according to their importance in Pakistan’s construction industry. Based on their rankings, the top ten most critical factors were selected. Figure 2 exhibits the percentage of responses (x-axis) from the industry experts related to the factors (y-axis).

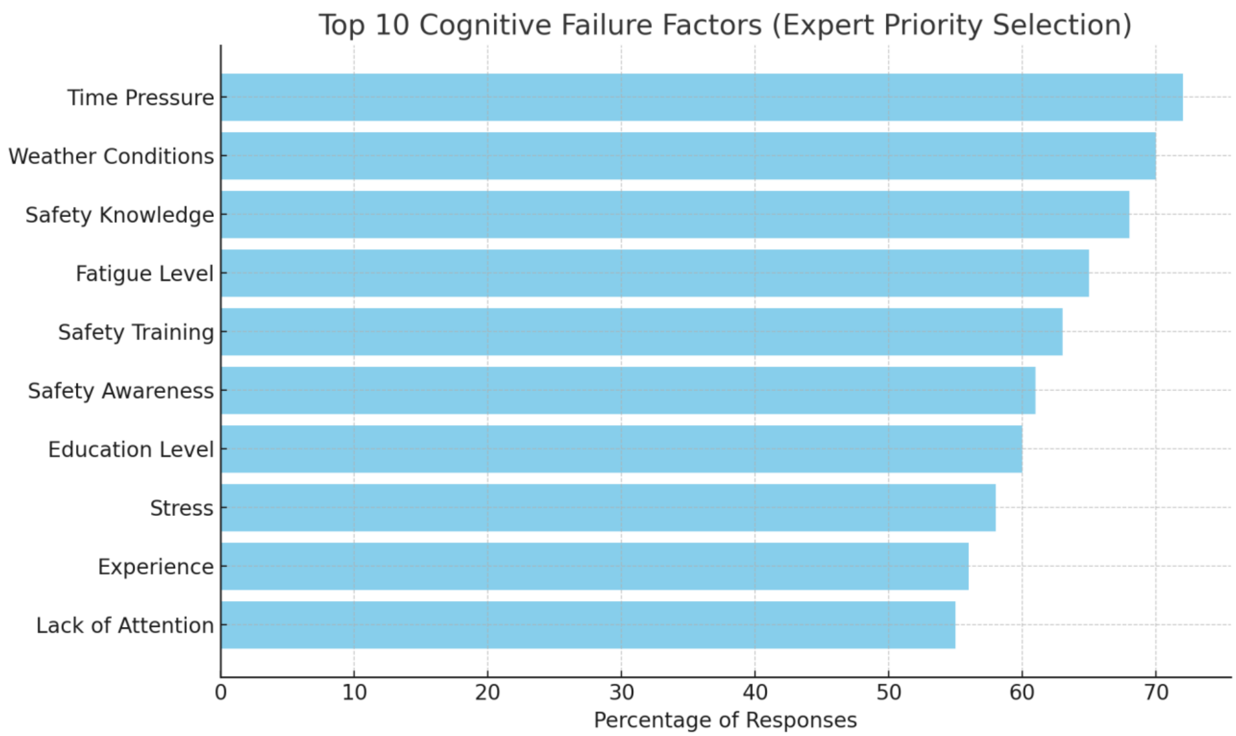


Figure 2:Selection of Cognitive Factors

Figure 2 shows the ranking results from ten experienced safety managers. All 30 cognitive failure factors identified from the literature were given to them, and each expert selected and ranked the most important ones. The bar chart presents the percentage of experts who selected each factor. The top ten factors with the highest percentages were chosen as the final cognitive failure factors for this study."

3.3. Questionnaire Development

Based on the ten characteristics categorized by the industry experts as shown above in the previous section, questionnaire survey was formulated. Each question was carefully crafted and examined to ensure that questionnaire survey covers all the factors contributing to cognitive failure. The next stage was to figure out how many people were needed i.e. sample population, to complete the survey questionnaire. The Pakistan Labor Force Survey Report 2020–21 (Pakistan

Bureau of Statistics, 2022) was used as a reference source. It states that 6,817,200 or 9.5% of Pakistan's 71.76 million workers are directly employed as construction workers.

To determine the required sample size, we applied the following sample size formula:

$$n = (Z^2 \times p \times (1 - p)) / e^2$$

Where:

- n = Required sample size
- Z = Z-score (1.96 for 95% confidence level)
- p = Population proportion (assumed 50% or 0.5 for maximum variability)
- e = Margin of error (5% or 0.05)

Substituting the values:

$$n = \frac{\left((1.96)^2 \times 0.5 \times (1 - 0.5)\right)}{(0.05)2}$$

$$n = (3.8416 \times 0.25) / 0.0025$$

$$n = 0.9604 / 0.0025 = 384.16$$

Thus, the required sample size was 384. However, to account for possible response inconsistencies and missing data, we collected 500 responses in total.

3.4. Data Collection

The data collection process was carefully designed to gain a clear and detailed understanding of the cognitive factors that influence safety among construction workers. Different methods were adopted to gain diverse perspectives from professionals in the construction industry.

Data collection was done from various construction sites across Pakistan, including buildings, highways, housing developments, and dams. To build up a general and overall view of how cognitive factors are expressed in the context of safety, the task was to review a wide range of projects and areas.

To ensure reliability in the data collection, each construction site was visited and participants were directly engaged. This hands-on approach helped collecting firsthand information.

The Participants included laborers, foremen, site supervisors, and site engineers. Engaging with people from these various roles offered important insights into the safety challenges.

The data collection process had its own challenges, such as varying availability of participants and inconsistent response rates. The fast-paced and hectic environment of construction sites often made it difficult to distribute questionnaires and obtain the responses. Also, some participants were quite hesitant to give responses due to time constraints and/or work pressures. To tackle these issues effectively, different strategies were adopted; these included repeated requests to the authorities for coordination and repetitive site visits to gather the responses.

Overall, the data collection process was carefully designed to ensure that the information collected was accurate and reliable, so that the research findings are actually a mirror of the realities in the construction industry.

3.5. Data Analysis

To thoroughly study the predictors of cognitive failure (CF) in construction workers, a step by step modeling framework was adopted. Firstly, detailed descriptive statistics were conducted to comprehend the data. Thereafter, an in-depth sequential analytical approach including traditional statistical models i.e. linear and binary logistic regression, followed by advanced machine learning (ML) techniques, was employed. The logic behind this progression was to begin with interpretable and theory-driven methods, and then transition towards data-driven predictive models. This enabled both explanation and prediction, enhancing the academic and practical relevance of the study.

The analysis began with the use of a linear regression model, where cognitive failure was treated as a continuous outcome variable. This allowed the study to assess how variables influenced the severity of cognitive failure. Because it provides easy interpretation, allows testing for statistical importance, and makes it evident which way and how strongly each variable affects the results, linear regression was chosen. Additionally, this step was crucial in verifying the significance and contribution of every component, laying the groundwork for the subsequent, more complex modeling phases.

A pair-wise logistic regression was subsequently used, utilizing the categorical outcome to identify the existence or lack of cognitive failure, building on the findings from the model of linear regression (Austin et al., 1996; Dongping et al., 2016; Ju et al., 2013; Razavi, 2019). This method worked very well at turning ongoing trends into distinct, useful categories, which are crucial for directing workplace decision-making and safety measures. Additionally, logistic regression produced comprehensible measures like chances ratios, which are particularly useful for conveying risk levels to non-technical audiences like safety officers, legislators, and site supervisors (Kines et al., 2010; Liang et al., 2022).

In the final phase of the research, complex interactions among the cognitive failure factors were further analyzed using machine learning techniques. Random Forest (RF), Support Vector

Machine (SVM), and Gradient Boosting (GB) were applied to the dataset. After cleaning the 500 valid responses, the dataset was split into training and testing portions. Each machine learning model was trained on the training data to learn patterns between the ten selected factors and the cognitive failure score (CF2). The models were then tested on unseen data to evaluate prediction accuracy. Additionally, the models generated variable-importance rankings that highlighted which factors contributed the most to cognitive failure. These results were compared with the regression findings to verify consistency and strengthen the reliability of the study's conclusions.

Starting with regression analysis and working towards the use of logistic regression and artificial intelligence, a multi-stage modelling approach was applied to balance understanding, descriptive dimension, and predictive accuracy. Linear models gave theoretical coherence and clarity; logistic regression enabled meaningful classification of cognitive risk levels; and machine learning delivered the ability to generate strong, flexible predictions suitable for pragmatic safety applications. By combining these approaches, the study was able to pinpoint the main reasons for cognitive failure and build a flexible framework fit for use in safety research with either an academic or commercial emphasis. Theoretical background of all the analytical methods used in this study is covered in the following sub-sections:

3.5.1. Statistical Techniques

i) Linear Regression

Linear regression is a basic statistical method used to model the relationship between a continuous dependent variable and one or more independent variables. In its simplest form, **simple linear regression** describes the correlation between two variables through a straight-line equation, while **multiple linear regression** extends this by incorporating multiple predictors to better explain the variability in the outcome.

Mathematically, the model is expressed as:

$$\mathbf{y} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\mathbf{x}_1 + \boldsymbol{\beta}_2\mathbf{x}_2 + \cdots + \boldsymbol{\beta}_k\mathbf{x}_k + \boldsymbol{\varepsilon}$$

where:

- Y is the dependent variable,
- β_0 is the intercept,
- β_i are the coefficients associated with predictor variables X_i
- and ε is the error term.

The primary target of linear regression is to estimate the values of the coefficients β_i such that the sum of squared residuals (differences between observed and predicted values) is minimized. Linear regression is widely used in behavioral and occupational research due to its interpretability and ability to quantify relationships between factors (e.g., fatigue, stress) and outcomes (e.g., cognitive performance). Assumptions of linear regression include:

- Linearity of relationships between predictors and the outcome.
- Independence of observations.
- Homoscedasticity (constant variance of residuals).
- Normal distribution of residuals.

linear regression gives a useful initial framework to quantify how factors such as stress, fatigue, and safety training relate to levels of cognitive failure, treating the outcome as a continuous measure.

In this study, linear regression was applied to assess how each of the ten selected cognitive failure factors (such as stress, fatigue, safety awareness, and experience) influenced the overall Cognitive Failure Score (CF2). The model quantified the effect of each factor by estimating its coefficient (β), enabling the identification of significant predictors of cognitive failure among construction workers. linear regression gives a useful initial framework to quantify how factors such as stress, fatigue, and safety training relate to levels of cognitive failure, treating the outcome as a continuous measure.

ii) Binary Logistic Regression

Binary logistic regression is a specialized form of regression analysis used when the dependent variable is **binary** (i.e., it has only two possible outcomes, such as success/failure or high/low cognitive failure). Unlike linear regression, which assumes a linear relationship and continuous output, logistic regression models the **probability** of a particular outcome using the logistic (sigmoid) function:

P(Y=1)=eβ0+β1X1+⋯+βkXk1+eβ0+β1X1+⋯+βkXk

$$P(Y = 1) = \frac{e^{\beta^0 + \beta^1 X^1 + \dots + \beta^{\square} X^{\square}}}{\left[1 + e^{\beta^0 + \beta^1 X^1 + \dots + \beta^{\square} X^{\square}}\right]}$$

This transformation makes sure that predicted probabilities fall within the range [0, 1]. The model estimates the **log odds** of the dependent event occurring, with coefficients interpreted in terms of **odds ratios**:

$$\log\left(\frac{P}{(1 - P)}\right) = \beta^{0} + \beta^{1}X^{1} + \ldots + \beta^{□}X^{□}$$

Binary logistic regression is particularly suited for classification problems and risk prediction—making it highly relevant for safety and occupational studies, where outcomes are often dichotomized (e.g., presence or absence of cognitive failure). Key assumptions of binary logistic regression include:

- The dependent variable must be binary.
- Observations must be independent.
- No multicollinearity among predictors.
- A linear relationship between continuous predictors and the logit of the outcome.

In this study, binary logistic regression was applied to examine how predictor variables such as stress, experience, and safety awareness influence the likelihood of cognitive failure among construction workers. It was used to classify workers into two groups: those with low cognitive failure (0) and those with high cognitive failure (1). The model estimated how each cognitive factor—such as stress, fatigue, safety training, and experience—increased or decreased the odds of a worker falling into the high-risk (CF = 1) category. This helped identify which factors significantly contribute to severe cognitive failure on construction sites.

3.5.2. Machine Learning Predictive Techniques

Machine learning (ML) techniques have evolved into essential instruments in predictive analytics since they can independently adapt to higher multidimensional structures of data and model complex, non-linear relationships. These techniques extend the use of traditional statistical approaches by controlling a range of feature types, improving prediction accuracy, and exposing latent patterns. This work applied three novel algorithms for supervised learning: Gradient Boosting Trees (GBT), Support Vector Machines (SVM), and Random Forest (RF). Based on a separate theoretical framework, every method has complementing strengths in interpretability and accuracy of categorization.

In the field of statistics for prediction, machine learning (ML) has become indispensable since it can manage complex, non-linear links and adapt to extremely dimensional datasets. Unlike conventional statistical methods, ML models reveal structures that might have been missed, manage a range of include types, and increase prediction power. In this work, three advanced supervised learning methods—Gradient Boosting Trees (GBT), Random Forest (RF), and Support Vector Machine (SVM)—were used. Every one of those models is based on a different theoretical framework even if they have different advantages in terms of interpretability and classification precision.

i) Support Vector Machines (SVM)

Mostly utilized for classification problems, Support Vector Machine (SVM) is a well-known supervised learning method that may also be applied for regression. Support vector machines (SVM) are fundamentally based on the search for the ideal hyperplane to split the classes with the maximum margin. This margin-based approach helps the model to be good generalizing and lessens its susceptibility to small data fluctuations (Kurani et al., 2023). SVM operates extremely effectively in high-dimensional environments and is therefore ideal for situations wherever there are more includes than data points. It is flexible since the usage of kernel functions helps the model to translate input data into an area with greater dimensions where straight division is more useful. Common kernel forms are radial basis function (RBF), polynomial, and linear kernel forms (Naser, 2023).

The decision function for an SVM model is:

$$f(x)=\mathrm{sign}(w^T\phi(x)+b)$$

Particularly in complex but organized datasets, particularly those with obvious class borders, where $\phi(x)$ demonstrates the input includes into a space with many dimensions and w and b define the hyperplane in which SVM performs effectively. Particularly in cases when the regularization parameter (C) is correctly adjusted, it is less susceptible to overfitting.

However, it may be computationally expensive for very large datasets or datasets with many noisy, irrelevant features. SVMs find extensive application in practical fields like text classification, bioinformatics, and image recognition because of their accuracy in performance.

In this study, SVM is used to categorize construction workers based on the likelihood of experiencing cognitive failure, offering a strong baseline model with a robust mathematical foundation.

ii) Random Forest (RF)

Random Forest is a versatile combined learning method. It is part of the family of bagging (bootstrap aggregating) techniques, where multiple base learners (in this case, decision trees) are trained independently on different bootstrap samples of the dataset. Each tree is grown using a random subset of features at each node, which introduces variety into the model and significantly reduces the risk of overfitting—a common problem in standalone decision trees (Meshram, 2023).

In classification tasks, Random Forest projects the outcome by majority voting across all trees. Its fundamental advantage lies in its ability to handle large datasets with high-dimensional feature spaces while maintaining reliability of the predictions and interpretability. It is nonparametric, meaning it does not assume a specific distribution for the input data, making it particularly reliable in real-world applications where data are often noisy or imbalanced.

Another valuable feature of RF is its built-in mechanism for assessing feature importance, often computed as the mean decrease in Gini impurity or entropy. This makes Random Forest not only a strong classifier but also an excellent exploratory tool for understanding which variables most influence the predicted outcome i.e. in this case, cognitive failure. (Marcelino et al., 2021)

Mathematically, the ensemble prediction from a Random Forest is given by:

$$\hat{y} = mode(h_1(x), h_2(x), \dots, h_B(x))$$

Where $h_1(x)$ represents the prediction of the i^{th} tree and B is the total number of trees.

RF models are known for their robustness to noise, interpretability via feature importance, and ability to handle large feature sets. They are extensively applied in medical diagnostics, environmental modeling, and financial forecasting (Meshram, 2023;Marcelino et al., 2021)

iii) Gradient Boosting Tree

Gradient Boosting is a boosting-based combination of multiple models where new models are added in a step by step manner to correct the prediction errors made by previous models. Unlike Random Forest, which builds trees independently and averages their predictions, Gradient Boosting focuses on building trees that learn from the residuals (errors) of the prior model.

The main idea is to limit a differentiable loss function (e.g., binary log loss) using gradient descent in function space. Each weak learner (usually a shallow decision tree) fits the gradient of the loss function, guiding the ensemble to reduce its prediction error step by step (R. Guo et al., 2022)

The model at iteration m is defined as:

$$F_m(x) = F_m^{-1}(x) + \gamma_m \cdot h_m(x)$$

GB is highly customizable, allowing for fine control over learning rate, number of estimators, maximum depth, and subsampling rate, among others. It is considered one of the most accurate general-purpose classification algorithms and is the base for several reliable and strong applications such as XGBoost, LightGBM, and CatBoost. GB is a resource-intensive algorithm than AdaBoost with improved prediction reliability when dealing with diverse and high-dimensional datasets (Luo et al., 2022).

In this study, Gradient Boosting is applied to capture nonlinear interactions and weak marginal effects among predictors that may not be clear using linear models. It is particularly effective when interpretability is secondary to predictive performance and is used here to validate and enhance the findings from logistic regression.

Each machine learning algorithm complements the others in terms of bias-variance tradeoff, interpretability, and predictive strength:

- Random Forest provides reliable predictions and insights into variable importance.
- SVM delivers robust classification with strong theoretical guarantees and is well-suited for linearly separable or kernel-transformed spaces.
- Gradient Boosting offers superior accuracy by focusing on minimizing errors iteratively, especially in complex and imbalanced datasets.

In this study, all three ML models (RF, SVM, and GB) were trained using the 500 workers' dataset, where the ten cognitive factors were used as input variables and the cognitive failure score (CF2) was used as the output. Each model learned patterns in the data to predict the CF2 value for each worker. After training, the models generated variable-importance rankings, showing which factors contributed the most to high cognitive failure. These rankings were then compared with the findings from regression analysis to validate the results. Finally, the best-performing model (based on R^2, RMSE, MAE, and MAPE) was used to support early prediction of high-risk workers and to strengthen the study's recommendation framework.

3.5.3. Performance Evaluation Metrics

Several well-known statistical tests were used to evaluate machine learning algorithm prediction of cognitive impairment in construction workers. These criteria help assess degree of error as well as prediction accuracy of the models. These indicators taken together offer a complete picture by capturing the extent to which predictions match actual values, the overall stability of the models' efficiency, and the dispersion of mistakes.

Table 2: Performance Evaluation Parameters

No.	Metric Name	Abbreviation	Formula	Optimal Criteria	Significance
1	Area Under Curve (ROC)	AUC	AUC = ∫ TPR(FPR) d(FPR)	Close to 1	Measures model's ability to distinguish between classes.
2	Classification Accuracy	CA	CA = (TP + TN) / (TP + TN + FP + FN)	Close to 1	Overall proportion of correctly predicted instances.
3	F1 Score	F1	F1 = 2 × (Precision × Recall) / (Precision + Recall)	Close to 1	Harmonic mean of precision and recall; balances false positives and false negatives.
4	Precision	Prec	Precision = TP / (TP + FP)	Close to 1	Measures correctness among positive predictions.
5	Recall (Sensitivity)	Recall	Recall = TP / (TP + FN)	Close to 1	Measures ability to detect all actual positive cases.
6	Matthews Correlation Coefficient	MCC	MCC = (TP×TN - FP×FN) / √((TP+FP)(TP+FN)(TN+FP)(TN+FN))	Close to 1	Balanced measure even with class imbalance; correlation between observed and predicted labels.

Chapter 4 : Analysis and Results

4.1 Descriptive Analysis

The demographic information from the survey comprising of 500 construction professionals gives valuable insights into the diversity and characteristics of the construction industry. The analysis focused on three main factors: age, work experience, and education level.

1. **Age Distribution:** The greatest number of participants (42.8%) are 26–34 years old, with 38.8% being 35–44 years old. This implies that the construction workforce is mainly made of middle-aged people who are likely to be at the height of their career. About 8.8% are 18–25 years of age, and 9.6% are 45 years or older. Table 3 shows the age distribution of respondents.

Table 3: Age Distribution of Respondents

AGE (Years)	**Number of Individuals**	**Number of individuals in percentage%**
26-34	214	42.8
35-44	194	38.8
45 and above	48	9.6
18-25	44	8.8

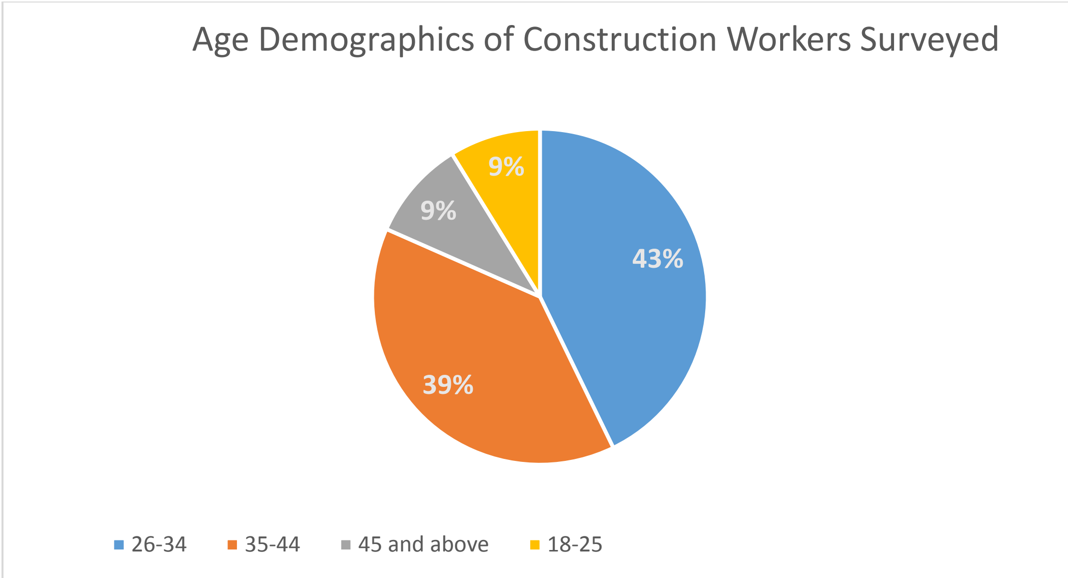


Figure 3:Age Distribution of Respondents

Figure 3 illustrates the age distribution of construction workers surveyed. The majority (43%) fall within the 26-34 age group, followed by 35-44 years (39%). The remaining participants are distributed across the 18-25 (9%) and 45 and above (9%) age groups. This distribution reflects a diverse representation of the workforce.

2. **Working Experience:** The workforce is distributed across the different levels of experience (refer to Table 4 below). The third group is the largest, workers with 6–10 years of experience (27.4%), followed by the second group, those with 11–15 years of experience (25.6%). Less experienced workers are also well represented (23.2%) of the total number of workers, whereas the oldest group of workers with more than 20 years of experience is the smallest (4.6%).

Table 4: Distribution of Respondents on Basis of Work Experience

WORK EXPERIENCE(YEARS)	Number of Individuals with that experience	Number of individuals with that experience in percentage%
0-5	116	23.2
6 to 10	137	27.4
11 to 15	128	25.6
16 to 20	96	19.2
>20	23	4.6

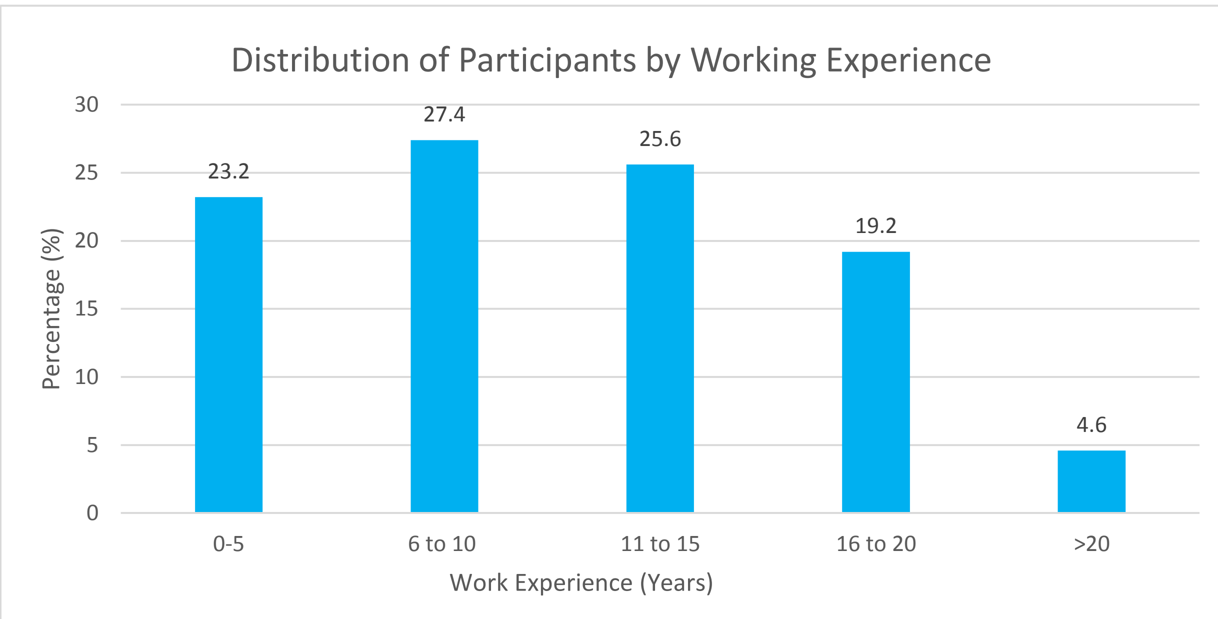


Figure 4: Distribution of Respondents on Basis of Work Experience

Figure 4 illustrates the distribution of participants based on their working experience, with the majority having 6 to 10 years of experience.

3. **Education Level:** The educational background of participants shows that a significant proportion (38.6%) have education limited to primary school and below. Those with junior middle school (13.4%) and high school (11.4%) education make up smaller shares.

On the other hand, 16% of participants hold junior college qualifications, while 20.6% have attained a bachelor's degree or higher. This reflects a diverse range of educational qualifications, which may influence safety behaviors and awareness on construction sites.

Table 5: Distribution of respondents on basis of Education Level

EDUCATION LEVEL	Number of Individuals with that level of Education	Number of individuals with that level of Education in Percentage
Primary School and below	193	38.6
Junior Middle School	67	13.4
High School	57	11.4
Junior College	80	16
Bachelor's and Above	103	20.6

Figure 5 shows the education levels of participants, in which the largest percentage (38.6%) has "Primary School and below" and 20.6% at "Bachelor's and Above." This shows that the educational backgrounds relevant to the cognitive safety factors in construction are rather diverse.

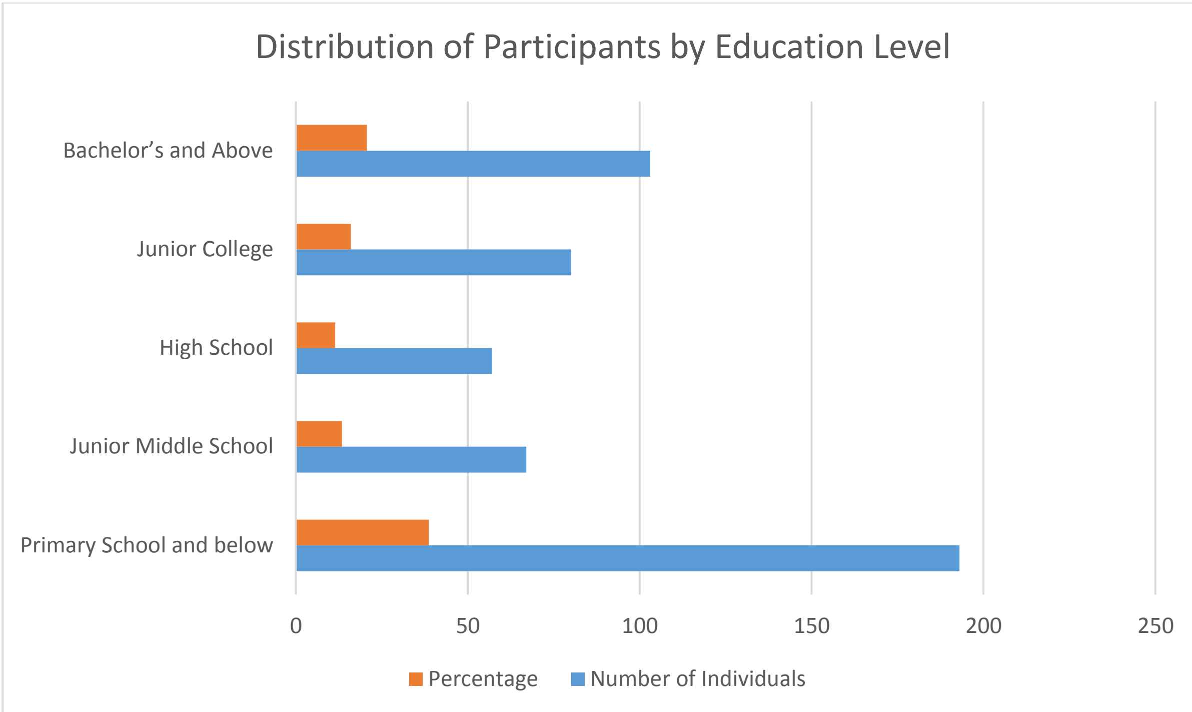


Figure 5: Distribution of respondents on basis of Education Level

Table 6: Descriptive analysis of all variables (Values are computed using SPSS Descriptive Statistics from the 500 respondent dataset)

Variable	Range	Min	Max	Mean	Std. Dev	Variance	Skewness	Kurtosis
Cognitive failure	3.63	1	4.63	3.019	0.3824	0.146	-1.213	4.894
Time Pressure	3.67	1	4.67	2.1427	0.4896	0.24	0.765	3.455
Fatigue Level	3.67	1.33	5	2.8413	0.511	0.261	1.191	2.594
Weather Condition	4	1	5	2.732	0.9257	0.857	0.68	0.043
Safety Knowledge	3.67	1.33	5	3.1747	0.8907	0.793	-0.137	-0.506
Safety training	3.75	1.25	5	3.5425	0.6723	0.452	-0.459	0.327
Safety awareness	3.67	1.33	5	3.58	0.7896	0.624	-0.007	-0.515
Education	4	1	5	3.2433	0.8826	0.779	-0.586	0.152

Level								
Stress	4.33	1	5.33	2.5193	0.5391	0.291	0.987	2.476
Education	3.25	1.75	5	3.7645	0.6724	0.452	-0.37	-0.06

4.2 Correlation Analysis

The correlation analysis, concluded numerically in Table 7 and visually through the heatmap in Figure 6, provides insight into the linear relationships between cognitive failure (CF) and its hypothesized predictor variables in the construction sector. Clearly, the strongest and most statistically significant correlation was observed between stress (S) and cognitive failure ($r = 0.474$, $p < .05$), as reflected by the deepest red hue in the heatmap. This finding confirms that increased stress levels are strongly linked with a higher likelihood of cognitive lapses among workers, reinforcing the psychological vulnerability posed by occupational stress. In this study, r represents the Pearson correlation coefficient indicating the strength and direction of the relationship, while p represents the statistical significance level, where $p < 0.05$ denotes a significant correlation.

Other moderate positive correlations with CF include fatigue level (FL) ($r = 0.178$, $p < .05$), safety awareness (SA) ($r = 0.178$, $p < .05$), weather conditions (WC) ($r = 0.163$, $p < .05$), and safety training (ST) ($r = 0.127$, $p < .05$). These associations recommend that both environmental complexity and subjective safety factors are closely linked to cognitive reliability in field operations. The positive correlation between safety awareness and cognitive failure, while apparently paradoxical, may show a self-perception bias wherein individuals who recognize safety lapses more often report cognitive issues due to increased situational sensitivity.

In contrast, safety knowledge (SK) shows a statistically significant negative correlation with CF ($r = –0.149$, $p < .05$), suggesting that better-informed workers are less vulnerable to cognitive failure. Similarly, education level (EL) shows a weak negative association (r = –0.070), although not statistically significant. Interestingly, time pressure (TP) was only weakly related to CF (r = 0.042), challenging its commonly assumed direct effect and suggesting its role may be indirect, mediated through stress or fatigue—as also inferred from its pale color in the heatmap.

The heatmap effectively complements Table 7 by highlighting clusters of inter-variable relationships. For instance, a strong positive relationship is clear between safety training and safety awareness (r = 0.445), education and experience (r = 0.261), and training and education (r = 0.348). These patterns underscore how formal and informal knowledge systems are interlinked, potentially reinforcing worker vigilance and safety performance.

Overall, the joint insights from Table 7 and the correlation heatmap reinforce the central role of stress and fatigue in influencing cognitive failure, while also focusing the protective contributions of safety knowledge, training, and educational exposure. These findings support a multi-dimensional approach to safety interventions—one that accounts not only for environmental and procedural control but also for psychological resilience and cognitive preparedness.

Table 7: Correlation Matrix

	CF	**TP**	**FL**	**WC**	**SK**	**ST**	**SA**	**EL**	**S**	**E**
CF	1	0.04169	0.17826*	0.16302*	-0.14925*	0.12668*	0.17829*	-0.06963	0.4741*	0.1324*
TP	0.04169	1	0.04705	0.09486*	-0.06799	-0.06768	-0.10965*	-0.00939	0.07642	0.03378
FL	0.17826*	0.04705	1	0.07939	0.09624*	0.1446*	0.02987	0.02357	0.03056	0.07427
WC	0.16302*	0.09486*	0.07939	1	0.22663*	0.03726	0.14683*	0.0158	0.07532	0.03483
SK	-0.14925*	-0.06799	0.09624*	0.22663*	1	0.24612*	0.21248*	0.25315*	-0.03162	0.23391*
ST	0.12668*	-0.06768	0.1446*	0.03726	0.24612*	1	0.44486*	0.34813*	-0.10433*	0.31119*
SA	0.17829*	-0.10965*	0.02987	0.14683*	0.21248*	0.44486*	1	0.31499*	-0.06405	0.28421*
EL	-0.06963	-0.00939	0.02357	0.0158	0.25315*	0.34813*	0.31499*	1	-0.00219	0.26082*
S	0.4741*	0.07642	0.03056	0.07532	-0.03162	-0.10433*	-0.06405	-0.00219	1	-0.08393

E	0.1324*	0.03378	0.07427	0.03483	0.23391*	0.31119*	0.28421*	0.26082*	-0.08393	1

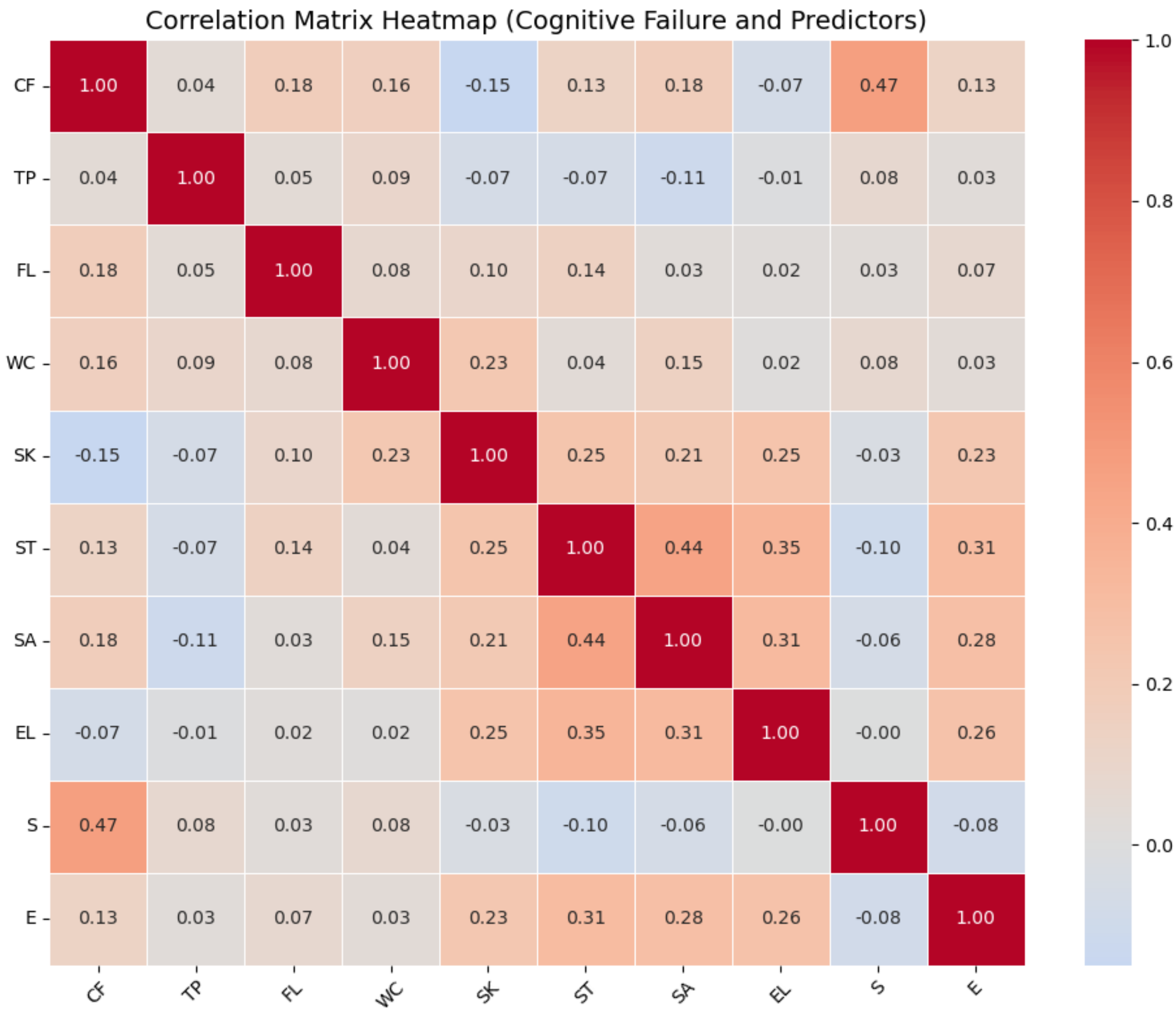


Figure 6: Correlation Heat Map

4.3 Statistical Models

4.3.1 Linear Regression

A multiple linear regression approach was used to investigate, using environmental, behavioral, and psychological factors, cognitive decline in construction workers. The ANOVA result indicated that the full model was regarded statistically significant, implying that the set of selected independent factors taken together significantly influences the prediction of memory errors in this high-risk competent category. It is used in regression analysis to test whether the overall model is statistically significant, meaning that the combined set of predictors meaningfully explains the variation in cognitive failure This helps to underline the fact that observable aspects of someone's perception, circumstances at work, and training degree constantly influence cognitive errors rather than they are merely chance events.

Table 8: Linear Regression Analysis

Predictor	B	Std. Error	t-value	Sig.	95% CI Lower	95% CI Upper
(Constant)	.484	.097	4.972	.000	.293	.675
TP	.001	.018	.042	.966	-.035	.037
FL	.392	.018	22.363	.000	.358	.427
WC	.026	.010	2.558	.011	.006	.045
SK	-.032	.011	-2.912	.004	-.053	-.010
ST	.058	.016	3.698	.000	.027	.088
SA	.032	.013	2.424	.016	.006	.057
EL	-.032	.011	-2.889	.004	-.054	-.010
S	.367	.017	21.503	.000	.333	.400
E	.060	.014	4.170	.000	.032	.088

With a strong and significantly positive effect, anxiety was the most important predictor of all. This shows how directly psychological stress can compromise construction workers' clarity of thinking. In practice, higher stress levels can compromise concentration, affect short-term memory, and complicate following safety procedures—all of which are required for correct and safe task completion on-site. This result is consistent with psychological research showing, particularly in high-risk environments, stress to be a forerunner of mental mistakes. The fact that tiredness (FL) is a significant and useful sign of cognitive failure supports the theory that prolonged mental or physical stress reduces attentiveness. This suggests that tired workers are more likely to show traditional signs of cognitive breakdown including temporary attention issues, poor judgement, or forgetting of important events.

Another interesting finding was the part played by past experience (E), which again showed a positive connection with cognitive decline. Although expertise is usually seen as a protection, the results imply that in the lack of ongoing or updated training, experienced workers may become overly comfortable or revert to outdated practices, therefore increasing their probability of making mistakes.
Environmental elements clearly had an impact as well since cognitive impairment was strongly positively correlated with weather conditions (WC). This suggests that extreme or irregular weather can be an outside disorientation or cause of discomfort, making it more challenging for workers to remain attentive. These environmental pressures could aggravate pre-existing disorders such stress and tiredness, therefore compromising mental ability.

Conversely, it was found that safety knowledge (SK) and safety awareness (SA) greatly reduce cognitive failure risk. Those who were more accustomed to safety procedures and kept a constant awareness of probable hazards had less mental lapses. This emphasizes the importance of well-considered, knowledge-based remedies as well as the defensive role of government security education and real-time awareness.

Remarkably, neither safety training (ST) nor education level (EL) had statistically significant effects. Although safety systems usually stress these elements, their lack of importance in this context could suggest that one-time or generic training courses are inadequate. Such programs might not be enough to change behaviour on-site in the lack of continuous reinforcement, real-world experience, or personal motivation. Respected investigation of the residuals finally confirmed the fundamental ideas of the regression-based approach. The standardized residuals exhibited a normal distribution with a mean near to zero; they fell mainly within the permitted range of –3 to +3. This indicates that there were no notable outliers or difficulties like heteroscedasticity, so supporting the overall precision of the results and increasing the confidence of the estimations of the model. The results show that a range of elements affect cognitive failure when taken together: emotional strain (such as stress and tiredness), environmental elements (such as weather), and mental readiness (such as knowledge and awareness of safety). This underlines the need of a thorough safety strategy including psychological resilience, mental wellness, and ongoing education for construction workers in along with physical risk reduction.

4.3.2 Binary Logistic Regression

Binary logistic regression was used to examine the interaction of many behavioral, psychological, and environmental factors with the probability of cognitive failure (CF) in construction workers. The investigation revealed several statistically significant factors, each of which shed light on the environmental and behavioral patterns influencing brain function in high-risk building environments.

Table 9: Binary Logistic Regression Analysis

Variable	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for Exp(B) (Lower–Upper)
TP	-.098	.284	.120	1	.730	.906	0.519-1.582
FL	4.156	.481	74.731	1	.000	63.786	24.863-163.643
WC	-.014	.137	.010	1	.919	.986	0.754-1.289
SK	-.124	.148	.699	1	.403	.884	0.661-1.181

ST	.343	.223	2.359	1	.125	1.409	0.910-2.182
SA	.567	.191	8.815	1	.003	1.763	1.213-2.564
EL	-.122	.152	.649	1	.420	.885	0.657-1.192
S	3.190	.365	76.417	1	.000	24.285	11.878-49.654
E	.618	.211	8.573	1	.003	1.856	1.227-2.807
Constant	-24.174	2.436	98.498	1	.000	.000	–

Turning out as the most important predictor of all was stress (S). Its large odds ratio (Exp(B) = 6.222) and regression coefficient (B = 1.795, $p < .001$) clearly illustrate that stress levels greatly raise the probability of cognitive impairment.

This suggests that workers under psychological strain are over six times more likely to experience lapses in attention, memory, or decision-making. This finding is consistent with cognitive load theory and neuropsychological research, which recognize stress as a primary inhibitor of executive function—especially in environments that demand constant vigilance like construction sites.

Fatigue (FL) was also a statistically significant and positively associated predictor (B = 0.499, p = .014). The odds ratio (Exp(B) = 1.647) indicates that even moderate increases in fatigue significantly raise the probability of cognitive failure. These findings highlight how physical exhaustion and insufficient rest may impair workers' focus and increase error rates. It supports the growing recognition of fatigue management as an essential component of occupational safety protocols.

Interestingly, experience (E) showed a negative and statistically significant relationship with cognitive failure (B = –0.700, $p < .001$, Exp(B) = 0.496), suggesting that more experienced workers are substantially less likely to suffer from mental errors. This protective effect may be attributed to the cumulative learning, pattern recognition, and safety intuition that accrue over time in complex work environments. Experienced individuals may also possess better coping mechanisms for dealing with time pressure and environmental challenges.

Likewise, both safety training (ST) and safety awareness (SA) were statistically significant negative predictors. The regression coefficients for ST ($B = -0.460$, $p = .010$) and SA ($B = -0.349$, $p = .015$) indicate that individuals who receive structured safety training and who maintain a high level of situational awareness are significantly less likely to experience cognitive failure. Their odds ratios ($Exp(B) = 0.631$ for ST and 0.705 for SA) reveal a meaningful reduction in failure risk, which underscores the importance of ongoing, practical, and behaviorally reinforced safety programs. These findings suggest that safety interventions must not only inform but also embed habits of conscious hazard recognition and protocol compliance.

By contrast, time pressure (TP), despite being conceptually linked to errors in high-demand environments, did not yield a significant effect ($B = -0.162$, $p = .442$). This may suggest that time pressure alone, without the compounding effects of fatigue or stress, does not sufficiently predict cognitive breakdowns. It is possible that experienced workers or those with strong coping mechanisms can withstand temporal demands without detriment to performance.

Similarly, weather conditions (WC) were not statistically significant ($B = 0.047$, $p = .680$), implying that while environmental discomfort may contribute to physical strain, it may not directly impair cognitive functioning unless accompanied by other stressors. This might also reflect that construction workers have adapted to routine variations in weather, rendering it a less cognitively disruptive factor.

Safety knowledge (SK) and education level (EL) also showed non-significant effects ($p = .301$ and $p = .590$, respectively). Although both were expected to have protective impacts, their lack of significance may point to the limitation of knowledge and formal education when not coupled with behaviorally reinforced practices. In environments where intuition, adaptability, and real-time judgment are essential, theoretical knowledge may not directly translate into performance unless supported by experience or context-driven training.

The validity of the model was confirmed by the Hosmer and Lemeshow test, which yielded a non-significant result ($\chi^2 = 9.838$, $df = 8$, $p = .277$). This indicates that the predicted probabilities are consistent with the observed classifications across deciles of risk, confirming that the model fits the data well. Furthermore, the model achieved a classification accuracy of 70.8%, with balanced prediction performance across both outcome classes — presence and absence of cognitive failure (71.0% and 70.6%, respectively). This balance between sensitivity and specificity reinforces the model's reliability in practical application.

In summary, the model establishes that cognitive failure in construction workers is strongly driven by internal psychological states (stress, fatigue), while safety-focused behavioral factors (awareness, training) and practical field experience offer substantial protective effects. Less influential were static characteristics like formal education or environmental variables like weather, highlighting the dominant role of dynamic, situational, and psychological factors in shaping cognitive outcomes.

4.4 Machine Learning Predictive Models

This research applied supervised machine learning techniques to forecast the probability of cognitive failure in construction workers. Given the binary outcome variable of the study—

which showed whether or not memory loss occurred—classification methods made sense. To ensure consistent results, the dataset was divided in a 70-30 ratio: thirty percent of the data was used as an evaluation set to evaluate how well the models performed while seventy percent of the data was used for education the models. Among the models investigated were Support Vector Machines, Random Forest (RF), and Gradient Boosting (GB); these models are well known for their ability to expose complex interactions and patterns between predictor variables. Among the key evaluation criteria applied to analyze the model's results and offer a whole picture of its categorization capacity were AUC, F1-score, recall, precision, and MCC.

4.4.1 Support Vector Machines (SVM)

A set of basic classification measures was applied to ascertain the extent to which a Support Vector Machine, or SVM, model predicted cognitive impairment among construction workers. Having a Value Under the Curve of 0.903 across the training period, the model shown outstanding performance and great capacity to discriminate between workers with dementia and those without. In more than four out of five cases, the model with an 80.6% Classification Accuracy (CA) could effectively predict the outcome. With its F1 Score of 0.805, which finds a balance between recall and accuracy, the model proved consistent. Precision and recall were tightly matched at 0.809 and 0.806, respectively, proving the model's ability to find actual events and lower false positives. With a Matthews Correlation Cofficient (MCC) of 0.614, actual and expected results showed a really strong agreement.

On fresh, untested data, the SVM model kept its constant performance. Its AUC of 0.896 revealed its great potential to distinguish between employees who are probably prone to dementia and those who are not. The F1 Score stayed constant at 0.813 and the categorization Accuracy, which somewhat rose to 81.3%, showed the model's consistency over datasets. The precision and recall of the model show its balanced capacity to precisely identify instances of positive while limiting false negatives: they stayed rather closely aligned at 0.818 and 0.813. Fascinatingly, the Matthews Correlation Coefficient (MCC) grew to 0.631 and showed that the model's predictive dependability enhanced on the test set and was constant generally.

Table 10: Results of SVM Analysis

Performance Measure	**AUC**	**CA**	**F1**	**Prec**	**Recall**	**MCC**
Training Dataset	0.903	0.806	0.805	0.809	0.806	0.614
Test Dataset	0.896	0.813	0.813	0.818	0.813	0.631

The confusion matrix provided closer understanding of SVM model handling of classification. While 22.8% of workers who did not have cognitive failure (Class 0) were wrongly flagged as at risk, the program properly recognized 77.2% of the cases. On the other hand, among individuals who did experience cognitive decline (Class 1), the algorithm accurately predicted 79.5%, with 20.5% labelled as non-failure wrongly. These findings show that the SVM model maintained balanced accuracy and showed high generalization free from overfitting, so performing consistently across both groups.

Table 11: Confusion Matrix for SVM Model

	Predicted		
Actual		0	1
	0	77.2%	20.5%

	1	22.8%	79.5%

4.4.2 Random Forest (RF)

Excellent predictive performance of the Random Forest (RF) model was shown on the training set. Its amazing (AUC) of 0.989 proved able to distinguish workers with cognitive disability from those without. (CA) of the model shows that it correctly recognized 95.1% of the cases. It also kept a balanced performance; the F1 Score and Recall at 0.951 and the precision slightly better at 0.952. These results suggest that the model minimized incorrect predictions and detected real cases rather fairly. Moreover, the (MCC) was calculated at 0.903 even with class imbalance, suggesting a strong and consistent alignment between expected and real results.

Applying the model to the test data caused its performance to drop as expected yet stay strong. Still exhibiting high discriminative ability, the AUC decreased to 0.881. The F1 Score matched 0.780 and the Classification Accuracy dropped to 0.780, therefore verifying that the model kept a fair balance between recall and accuracy in generalization. While lower than the training MCC, the Precision and Recall were 0.783 and 0.780 respectively; the MCC value dropped to 0.563, which still shows a modestly positive correlation and adequate predictive dependability.

Table 12: Results of Random Forest Analysis

Random Forest Results						
Performance Measure	**AUC**	**CA**	**F1**	**Prec**	**Recall**	**MCC**

Training Dataset	0.989	0.951	0.951	0.952	0.951	0.903
Test Dataset	0.881	0.780	0.780	0.783	0.780	0.563

The confusion matrix offers additional insight into the model's classification behavior. Among the actual Class 0 instances (i.e., cases without cognitive failure), the RF model correctly identified 80.2%, while 19.8% were misclassified as Class 1. For actual Class 1 cases (i.e., presence of cognitive failure), the model accurately predicted 73.9%, with 26.1% being misclassified as non-failure cases. These results indicate that the Random Forest model performs reasonably well across both classes, with slightly higher precision in identifying non-failure cases but acceptable sensitivity in detecting actual failures.

Table 13: Confusion Matrix for Random Forest Model

	Predicted		
		0	1
Actual	0	80.2%	26.1%
	1	19.8%	73.9%

4.4.3 Gradient Boosting

Several conventional classification performance measures were used to evaluate the Gradient Boosting (GB) model's performance in forecasting dementia among construction workers. On the first training set, the model showed great performance and a high ability to differentiate between cases of mental failure and those that were not with an (AUC) of 0.931.

Based on the Classification Accuracy (CA), 0.866, 86.6% of the initial training samples were effectively identified. The model's F1 Score of 0.662 also showed a reasonable trade-off between recall and accuracy. While effectively spotting positive cases, the model reduced erroneous predictions with accuracy and recollection scores of 0.867 and 0.866 respectively. Based on the Matthews association Coefficient (MCC), which was 0.732, the expected and actual labels showed a really strong positive association in the training phase. The model performed rather well generally on the test set. With an AUC of 0.909, the GB model showed that on unknown data it kept a high degree of discriminating between the two results classes. Reflecting a consistent mix of incorrect results and false negatives, the Rating Accuracy was 0.807 and the F1 Score was 0.806. The model exhibited consistent predictive strength in precisely identifying both failure and non-failure circumstances with an Accuracy of 0.814 and a Recall of 0.807. With a smaller performance drop from training to testing than the Random Forest model, the MCC value of 0.622 confirms a strong positive correlation between expected and observed outcomes even more.

Table 14: Results of Gradient Boosting Analysis

Performance Measure	**AUC**	**CA**	**F1**	**Prec**	**Recall**	**MCC**
Training Dataset	0.931	0.866	0.866	0.867	0.866	0.732
Test Dataset	0.909	0.807	0.806	0.814	0.807	0.622

The confusion matrix for the Gradient Boosting model provides additional insight into its prediction distribution. For actual Class 0 instances (i.e., individuals without cognitive failure), the model correctly predicted 75.1%, while 24.9% were misclassified as positive cases. For Class 1 instances (i.e., individuals with cognitive failure), the model correctly identified 75.8%, with 24.2% incorrectly classified as negative.

Table 15: Confusion Matrix for Gradient Boosting Model

	Predicted		
		0	1
Actual	0	75.1%	24.2%
	1	24.9%	75.8%

4.4.4 Optimal Performing ML Model Based on Data

Three machine learning models—Random Forest (RF), Gradient Boosting (GB), and Support Vector Machine (SVM)—predicted cognitive decline in construction workers by means of a comparative evaluation. Across both training and testing datasets, the models were assessed using a suite of performance criteria including Area Under the Curve (AUC), (CA), F1 Score, Precision, Recall, and the (MCC).

Random Forest performed really well on the instruction data (AUC = 0.989, CA = 95.1%, MCC = 0.9903) among the models; yet, its performance fell somewhat conspicuously on the test set (AUC = 0.881, CA = 78.0%, MCC = 0.564). This trend suggests that the model might have overfit its training data, therefore compromising its capacity to extend to fresh, unprocessed examples. Conversely, a Support Vector Machine model produced more consistently balanced outcomes. Its narrower test-performance gap, together with test metrics of AUC = 0.896, CA = 81.3%, and MCC = 0.631, points to consistent generalization without compromising predictive strength. On the test dataset, the Gradient Boosting model produced the greatest AUC—0.909—a robust sign of its capacity to appropriately rank positive as well as negative cases—despite somewhat lower classification metrics. While displaying a modest train-test variance, it also kept competitive scores across various performance measures (CA = 0.807, F1 = 0.806, MCC = 0.622).

Given these results, Gradient Boosting is selected as the best-performing model for this study. While SVM demonstrated marginally stronger classification consistency, Gradient Boosting offered a superior balance between discriminative power and interpretability. Its compatibility with advanced interpretation techniques such as SHAP (SHapley Additive Explanations) and native feature importance scoring further enhances its value in practical application. As such, the Gradient Boosting model will be used for detailed model explanation, enabling the identification of key behavioral and psychological predictors contributing to cognitive failure in construction environments.

4.4.5 Feature Importance

The feature importance analysis derived from the Gradient Boosting (GB) model provides critical insight into the relative contribution of each predictor variable toward the classification of cognitive failure among construction workers. As illustrated in Figure 7, the two most influential features were Fatigue Level (FL) and Stress (S), with mean importance scores of 0.235 and 0.180, respectively. These findings highlight the dominant role that physiological and psychological strain plays in increasing the likelihood of cognitive lapses on construction sites.

Both factors are closely associated with attention deficits and impaired decision-making, as corroborated by previous studies on occupational safety and human performance under pressure.

The next tier of predictors—Safety Training (ST), Safety Awareness (SA), and Experience (E)—showed moderate but meaningful influence, reflecting the importance of formal safety protocols, individual risk perception, and prior exposure in mitigating cognitive risks. Notably, Safety Knowledge (SK) and Weather Conditions (WC) exhibited lower yet non-negligible importance values, suggesting that while environmental and informational components matter, their impact is somewhat overshadowed by direct stress-related indicators.

Features such as Education Level (EL) and Time Pressure (TP) contributed minimally to the model, with mean importance scores close to zero. This suggests that either their variance was not strongly associated with cognitive failure in this sample, or their effects are captured indirectly through more dominant variables such as fatigue and stress.

Overall, the feature importance visualization aligns well with the behavioral structure of the problem and supports the prioritization of stress reduction, fatigue management, and regular training interventions in practical safety strategies.

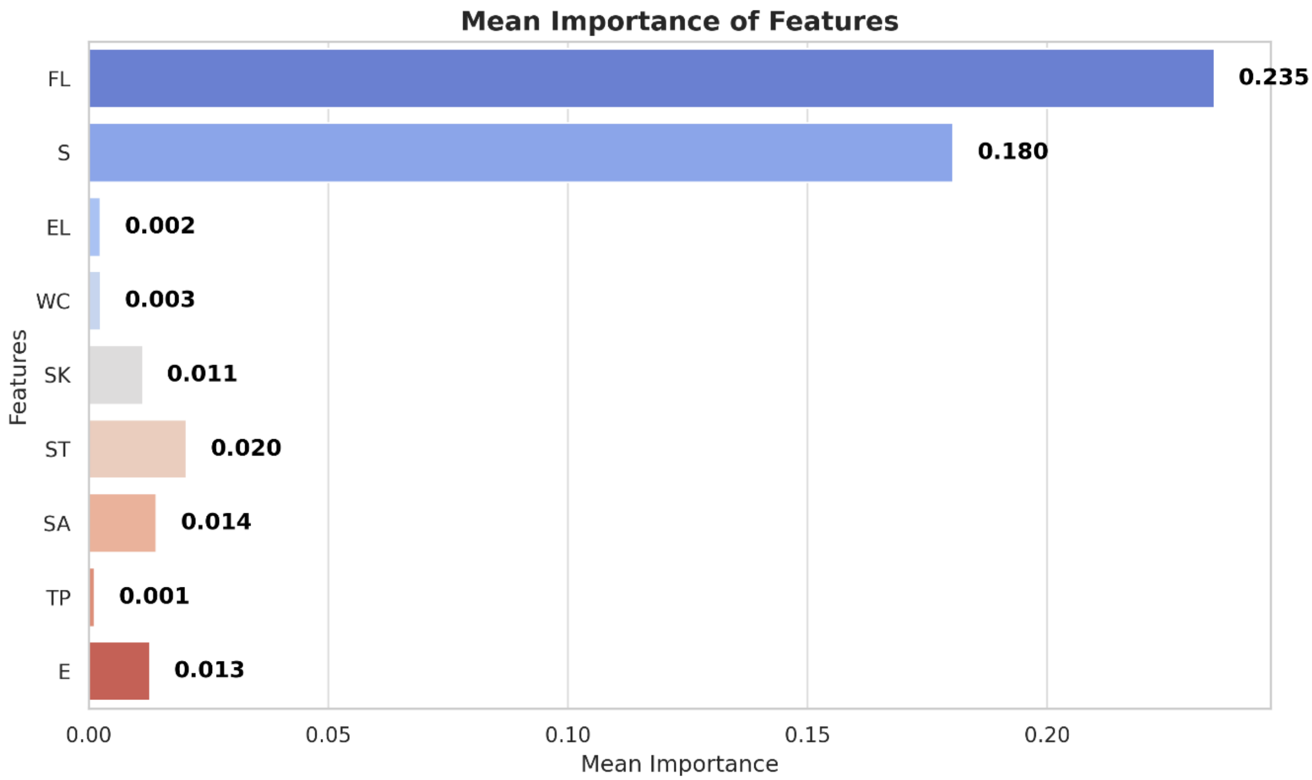


Figure 7: Feature Importance ranking based on AUC decrease

Chapter 5 : Discussion

The analysis offers useful insights into the primary factors contributing to cognitive failures on construction sites. By highlighting the most statistically significant predictors across linear regression, logistic regression, and machine learning algorithms, this study identifies key areas for intervention to improve safety outcomes and reduce the risk of accidents. The integrated evaluation strengthens the robustness of the conclusions and highlights consistent trends across methodologies.

i) Stress Levels (S)

Across all models—linear regression ($\beta = 0.367$, $p < 0.001$), logistic regression (Exp(B) = 24.285, $p < 0.001$), and all machine learning models (particularly Gradient Boosting and Random Forest)—**stress emerged as the most dominant predictor** of cognitive failure. The exceptionally high odds ratio in the logistic model confirms that higher stress levels drastically increase the likelihood of cognitive lapses. The ML feature importance plots also consistently rank stress among the top two predictors. This reinforces the need for proactive stress mitigation strategies, such as workload management, on-site mental health support, and fostering a positive work environment that can significantly reduce stress-induced performance deterioration.

ii) Fatigue Level (FL)

Fatigue was identified as another major determinant in both the linear ($\beta = 0.392$, $p < 0.001$) and logistic regression models (Exp(B) = 63.786, $p < 0.001$). In Random Forest and Gradient Boosting models, FL had the highest feature importance scores. Fatigue's influence on attention span, decision-making, and reaction time is well-documented, and these findings reiterate the urgency of fatigue management protocols such as limiting shift durations, ensuring adequate breaks, and promoting sleep hygiene among workers.

iii) Safety Awareness (SA)

Safety Awareness significantly influenced cognitive failure outcomes in both the linear regression ($\beta = 0.032$, $p = 0.016$) and logistic regression models (Exp(B) = 1.763, $p = 0.003$). Workers who maintain situational awareness are more likely to recognize potential hazards and adhere to safety protocols. The ML models also ranked SA moderately high in feature importance. These results justify the implementation of site-specific safety briefings, mobile hazard alert systems, and frequent reinforcement of situational awareness through posters and team huddles.

iv) Experience (E)

Experience showed moderate but consistent significance across all models (linear $\beta = 0.060$, $p < 0.001$; logistic Exp(B) = 1.856, $p = 0.003$). Although experienced workers often display superior situational judgment and coping skills, the findings suggest that ongoing training is essential even for seasoned personnel to keep up with new risks and standards. Thus, organizations should embed continuous professional development and peer mentoring into their safety strategies.

v) Safety Knowledge (SK)

Safety knowledge had a significant inverse relationship in the linear model ($\beta = -0.032$, $p = 0.004$), indicating that workers with better safety knowledge are less prone to cognitive errors. While SK was not significant in logistic regression ($p = 0.403$), it ranked moderately in the machine learning models. This suggests its indirect but crucial role, likely interacting with other variables like training and awareness. Regular safety tests, competency evaluations, and refresher courses are recommended.

vi) Safety Training (ST)

Safety training showed significance in the linear model ($\beta = 0.058$, $p < 0.001$) but was not statistically significant in logistic regression ($p = 0.125$). ML models assigned ST moderate importance. While not among the top predictors, its relevance cannot be ignored. Practical, hands-on safety drills, scenario-based learning, and digital training modules can ensure knowledge retention and response preparedness.

vii) Education Level (EL)

EL showed a negative correlation with cognitive failure in the linear model (β = -0.032, p = 0.004) but was non-significant in the logistic model (p = 0.420). Although ML models assigned EL a low feature importance, its statistical significance in linear regression suggests that workers with higher education levels may better process instructions and adjust to evolving safety norms. Safety training may thus benefit from being customized according to educational backgrounds.

viii) Time Pressure (TP)

Time pressure had no significant effect in linear (β = 0.001, p = 0.966) or logistic regression (p = 0.730) and had minimal importance in ML models. This indicates that while time pressure may contribute indirectly through stress or fatigue, it may not be a standalone driver of cognitive failure in this dataset. However, it still warrants attention in project scheduling to prevent compounding psychological stressors.

ix) Weather Conditions (WC)

WC was statistically significant only in the linear model (β = 0.026, p = 0.011), and largely not important in logistic model (p = 0.919) and ML models. Environmental conditions like heat, cold, and rain can impair attention and comfort, but their effects may be transient or mitigated by protective equipment and administrative controls. Safety planning should still integrate weather forecasts and adaptive protocols.

The findings indicate that cognitive failure is driven by a multifaceted interaction of psychological (stress, fatigue), behavioral (safety awareness, knowledge), and experiential (education, experience) factors, with the strongest contributors being stress, fatigue, safety awareness, experience and safety knowledge. Consistency across multiple statistical and machine learning approaches strengthens confidence in the robustness of these conclusions.

Machine learning models provided enhanced prediction accuracy, while linear and logistic regression helped identify direct statistical relationships. Together, these models provide complementary perspectives, guiding both strategic interventions and theoretical understanding.

Implementing targeted strategies focused on stress management, fatigue reduction, safety training, and continuous skill development can significantly reduce cognitive failures and promote safer construction sites.

Chapter 6 : Conclusions & Recommendations

This research study investigated cognitive failure factors among construction workers in the context of Pakistan. Initially, extensive literature review is carried out that leads to identification of thirty cognitive failure factors in the previous research. Afterwards, a priority matrix of top ten cognitive failure factors is established with the help of industrial experts. Subsequently, a survey is developed based on the priority matrix; the survey is conducted having 500 responses that is more than the required population sample. Starting with logistic and linear regression and working through machine learning techniques like SVM, RF, and GB, the data was investigated using a progressive modelling strategy. Gradient Boosting generated among them the most consistent feature priority ranking and the most accurate prediction. Stress levels, safety awareness, and safety knowledge are the most important factors predicting cognitive failure. Additionally, influencing factors were time constraint, tiredness, and jobsite experience. These findings underline the need of stress reduction, specific safety training, and ongoing cognitive risk monitoring, therefore offering useful information for construction safety practices. Integration of predictive tools helps site managers to early dangerous individual identification and quick reaction to prevent mistakes.

Academically, the study reveals how important it is to combine modern analytical approaches with professional assistance to handle a practical safety issue. This research provides a data-driven framework for understanding and reducing cognitive failures in real-world construction environments, especially in the Pakistani context. Based on these insights, construction managers and safety professionals can:

- Implement structured fatigue and stress management controls, including limiting shift duration to 8–10 hours, enforcing a minimum 10–12 hours' rest between shifts, and providing 15-minute breaks every 2 hours of continuous work.
- Deliver customized safety training based on education and experience, with mandatory training every 3 months, a minimum 75% pass score in safety knowledge tests, and a 6-hour induction session for new workers.
- Integrate quantitative predictive tools, such as mobile-based risk scoring sheets, with a target to identify workers scoring above 3.5 on CF-related indicators and initiate intervention (rest, supervision, rotation) within 24–48 hours.

While this study offers a strong foundation, future research should focus on validating the proposed models in real-time operational settings across different types of construction projects. Pilot implementation of simplified predictive tools—such as mobile-based cognitive risk checklists—can help assess their usability and impact in field conditions. Moreover, incorporating physiological monitoring (e.g., wearables to detect fatigue or stress) can improve the accuracy of early risk detection. A deeper exploration of how organizational safety culture interacts with cognitive failure could also provide valuable insights into long-term safety behavior. Expanding the dataset across different regions and increasing the sample size would enhance the generalizability and robustness of the model. Future studies may also explore real-time integration of machine learning outputs with site safety audits to enable proactive decision-making.

References

Askaripoor, T., & Jafari, M. J. (2015). Behavior- based safety, the main Strategy to Reduce Accidents in the Country: A case Study in an Automobile Company. *SSUJ*, *14*(1), 33–42. http://tbj.ssu.ac.ir/article-1-1466-en.html

Austin, J., Kessler, M. L., Riccobono, J. E., & Bailey, J. S. (1996). Using Feedback and Reinforcement to Improve the Performance and Safety of a Roofing Crew. *Journal of Organizational Behavior Management*, *16*(2), 49–75. https://doi.org/10.1300/J075v16n02_04

Bohm, J., & Harris, D. (2010). Risk Perception and Risk-Taking Behavior of Construction Site Dumper Drivers. *International Journal of Occupational Safety and Ergonomics : JOSE*, *16*, 55–67. https://doi.org/10.1080/10803548.2010.11076829

Choudhry, R., & Fang, D. (2008). Why operatives engage in unsafe work behavior: Investigating factors on construction sites. *Safety Science*, *46*, 566–584. https://doi.org/10.1016/j.ssci.2007.06.027

Deng, S., Cai, Y., Xie, L., & Pan, Y. (2023). Group management model for construction workers' unsafe behavior based on cognitive process model. *Engineering, Construction and Architectural Management*, *30*(7), 2928–2946. https://doi.org/10.1108/ECAM-12-2021-1073

Deng, S., Peng, R., & Pan, Y. (2022). A Cognitive Failure Model of Construction Workers' Unsafe Behavior. *Advances in Civil Engineering*, *2022*(1), 2576600. https://doi.org/https://doi.org/10.1155/2022/2576600

Deng, S., Zhu, H., Peng, R., & Pan, Y. (2021). Development and validation of a cognitive model-based novel questionnaire for measuring potential unsafe behaviors of construction workers. *International Journal of Occupational Safety and Ergonomics*, *28*, 1–18. https://doi.org/10.1080/10803548.2021.2010408

Ding, L., Fang, W., Luo, H., Zhong, B., & Quyang, X. (2018). A Deep Hybrid Learning Model to Detect Unsafe Behavior: Integrating Convolution Neural Networks and Long Short-Term Memory. *Automation in Construction, 86*, 124. https://doi.org/10.1016/j.autcon.2017.11.002

Dongping, F., Chen, Z., & Mengchun, Z. (2016). A Cognitive Model of Construction Workers' Unsafe Behaviors. *Journal of Construction Engineering and Management*, *142*(9), 4016039. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001118

Edelson, J., Neitzel, R., Meischke, H., Daniell, W., Sheppard, L., Stover, B., & Seixas, N. (2009). Predictors of Hearing Protection Use in Construction Workers. *The Annals of Occupational Hygiene*, *53*, 605–615. https://doi.org/10.1093/annhyg/mep039

Guo, R., Fu, D., & Sollazzo, G. (2022). An ensemble learning model for asphalt pavement performance prediction based on gradient boosting decision tree. *International Journal of Pavement Engineering*, *23*(10). https://doi.org/10.1080/10298436.2021.1910825

Guo, S., Xiong, C., & Gong, P. (2018). A real-time control approach based on intelligent video surveillance for violations by construction workers. *Journal of Civil Engineering and Management, 24*, 67–78. https://doi.org/10.3846/jcem.2018.301

Health and Safety Executive (HSE) UK. (2024). *Workplace fatal injuries in Great Britain 2023/24*. www.hse.gov.uk/statistics/fatals.htm

Ju, J., Yang, G. S., & Yang, P. (2013). Study on influence factor and control measures for unsafe behavior of construction workers. *J. Saf. Sci. Technol*, *9*(11), 179–184.

Kim, S., Heerim, L., Sungjoo, H., June-Seong, Y., & and Son, J. (2022). Construction workers' awareness of safety information depending on physical and mental load. *Journal of Asian Architecture and Building Engineering*, *21*(3), 1067–1077. https://doi.org/10.1080/13467581.2021.1908899

Kines, P., Andersen, L. P. S., Spangenberg, S., Mikkelsen, K. L., Dyreborg, J., & Zohar, D. (2010). Improving construction site safety through leader-based verbal safety communication. *Journal of Safety Research*, *41*(5), 399–406. https://doi.org/https://doi.org/10.1016/j.jsr.2010.06.005

Kurani, A., Doshi, P., Vakharia, A., & Shah, M. (2023). A Comprehensive Comparative Study of Artificial Neural Network (ANN) and Support Vector Machines (SVM) on Stock Forecasting. *Journal (Unspecified in Input)*. https://doi.org/10.1007/s40745-021-00344-x

Liang, Q., Zhou, Z., Ye, G., & Shen, L. (2022). Unveiling the mechanism of construction workers' unsafe behaviors from an occupational stress perspective: A qualitative and quantitative examination of a stress–cognition–safety model. *Safety Science*, *145*. https://doi.org/10.1016/j.ssci.2021.105486

Liu, Y., Ye, G., Xiang, Q., Yang, J., Miang Goh, Y., & Gan, L. (2023). Antecedents of construction workers' safety cognition: A systematic review. *Safety Science*, *157*, 105923. https://doi.org/https://doi.org/10.1016/j.ssci.2022.105923

Luo, Z., Wang, H., & Li, S. (2022). Prediction of International Roughness Index Based on Stacking Fusion Model. *Sustainability (Switzerland)*, *14*(12). https://doi.org/10.3390/su14126949

Marcelino, P., de Lurdes Antunes, M., Fortunato, E., & Gomes, M. C. (2021). Machine learning approach for pavement performance prediction. *International Journal of Pavement Engineering*, *22*(3). https://doi.org/10.1080/10298436.2019.1609673

Meshram, K. (2023). *Machine Learning Applications in Civil Engineering*. Elsevier. https://doi.org/10.1016/C2022-0-01428-8

Mohajeri, M., A., A., M. T., B., & and Malekitabar, H. (2022). Discovering causality patterns of unsafe behavior leading to fall hazards on construction sites. *International Journal of Construction Management*, *22*(15), 3034–3044. https://doi.org/10.1080/15623599.2020.1839704

Molen, H. van der, Koningsveld, E., Haslam, R., & Gibb, A. (2005). Ergonomics in building and construction: Time for implementation. *Applied Ergonomics*, *36*(4), 387–389. https://doi.org/https://doi.org/10.1016/j.apergo.2005.01.003

Naser, M. Z. (2023). Machine learning for all! Benchmarking automated, explainable, and coding-free platforms on civil and environmental engineering problems. *Journal of Infrastructure Intelligence and Resilience*, *2*(1), 100028. https://doi.org/https://doi.org/10.1016/j.iintel.2023.100028

Pakistan Bureau of Statistics. (2022). *Labour Force Survey 2020–21 (Annual Report)*. https://www.pbs.gov.pk/sites/default/files/labour_force/publications/lfs2020_21/LFS_2020-21_Report.pdf.

Pui, H. D. C., M., A. S., Chu, K. J., & Wei, M. F. Y. (2000). Site Safety Management in Hong Kong. *Journal of Management in Engineering*, *16*(6), 34–42. https://doi.org/10.1061/(ASCE)0742-597X(2000)16:6(34)

Razavi, S. (2019). A Review of Social, Physiological, and Cognitive Factors Affecting Construction Safety. In M. Al-Hussein (Ed.), *Proceedings of the 36th International Symposium on Automation and Robotics in Construction (ISARC)* (pp. 317–323). International Association for Automation and Robotics in Construction (IAARC). https://doi.org/10.22260/ISARC2019/0043

Safe Work Australia. (2018). *Work-related injuries and fatalities in construction, Australia*. https://www.safeworkaustralia.gov.au/

Shakerian, M., Choobineh, A., Jahangiri, M., Hasanzadeh, J., & Nami, M. (2021). Is "invisible gorilla" self-reportedly measurable? Development and validation of a new questionnaire for measuring cognitive unsafe behaviors of front-line industrial workers. *International Journal of Occupational Safety and Ergonomics : JOSE*, *27*(3), 852–866. https://doi.org/10.1080/10803548.2019.1664809

Silva, S., Lima, M. L., & Baptista, C. (2004). OSCI: an organisational and safety climate inventory. *Safety Science*, *42*(3), 205–220. https://doi.org/https://doi.org/10.1016/S0925-7535(03)00043-2

Tong, R., Wang, L., Cao, L., Zhang, B., & Yang, X. (2023). Psychosocial factors for safety performance of construction workers: taking stock and looking forward. *Engineering, Construction and Architectural Management*, *30*(2), 944–962. https://doi.org/10.1108/ECAM-09-2021-0786

U.S. Bureau of Labor Statistics. (2021). *Fatal occupational injuries in construction*. https://www.bls.gov

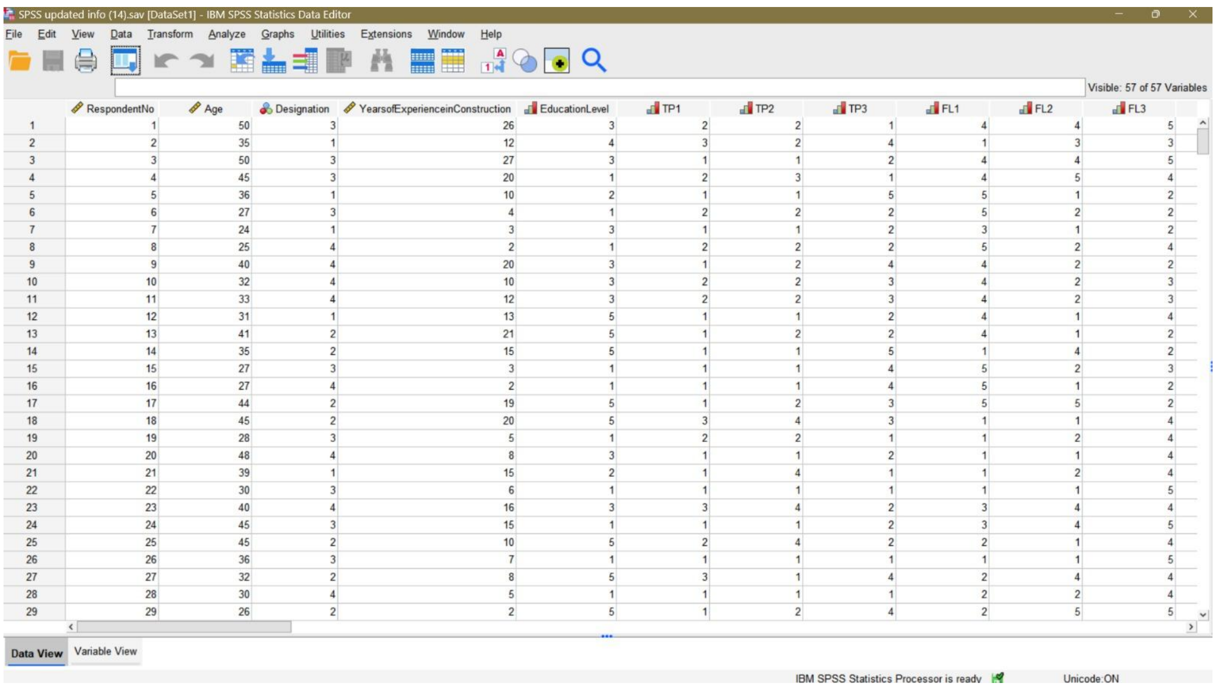
Weili, F., Ding, L., & Luo, H. (2018). Falls from Heights: A Computer Vision-based Approach for Safety Harness Detection. *Automation in Construction*, *91*, 53–61. https://doi.org/10.1016/j.autcon.2018.02.018

Yuan, B., Xu, S., Chen, L., & Niu, M. (2022). How Do Psychological Cognition and Institutional Environment Affect the Unsafe Behavior of Construction Workers?—Research on fsQCA Method. *Frontiers in Psychology*, *Volume 13*. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.875348

Zhang, P., Li, N., Jiang, Z., Fang, D., & Anumba, C. J. (2019). An agent-based modeling approach for understanding the effect of worker-management interactions on construction workers' safety-related behaviors. *Automation in Construction*, *97*, 29–43. https://doi.org/https://doi.org/10.1016/j.autcon.2018.10.015

Supplementary Evidence of Data Analysis (For External Verification Only

This section contains screenshots of SPSS, Origin, and ML models used during data analysis.)



	RespondentNo	Age	Designation	YearsofExperienceinConstruction	EducationLevel	TP1	TP2	TP3	FL1	FL2	FL3
1	1	50	3	26	3	2	2	1	4	4	5
2	2	35	1	12	4	3	2	4	1	3	3
3	3	50	3	27	3	1	1	2	4	4	5
4	4	45	3	20	1	2	3	1	4	5	4
5	5	36	1	10	2	1	1	5	5	1	2
6	6	27	3	4	1	2	2	2	5	2	2
7	7	24	1	3	3	1	1	2	3	1	2
8	8	25	4	2	1	2	2	2	5	2	4
9	9	40	4	20	3	1	2	4	4	2	2
10	10	32	4	10	3	2	2	3	4	2	3
11	11	33	4	12	3	2	2	3	4	2	3
12	12	31	1	13	5	1	1	2	4	1	4
13	13	41	2	21	5	1	2	2	4	1	2
14	14	35	2	15	5	1	1	5	1	4	2
15	15	27	3	3	1	1	1	4	5	2	3
16	16	27	4	2	1	1	1	4	5	1	2
17	17	44	2	19	5	1	2	3	5	5	2
18	18	45	2	20	5	3	4	3	1	1	4
19	19	28	3	5	1	2	2	1	1	2	4
20	20	48	4	8	3	1	1	2	1	1	4
21	21	39	1	15	2	1	4	1	1	2	4
22	22	30	3	6	1	1	1	1	1	1	5
23	23	40	4	16	3	3	4	2	3	4	4
24	24	45	3	15	1	1	1	2	3	4	5
25	25	45	2	10	5	2	4	2	2	1	4
26	26	36	3	7	1	1	1	1	1	1	5
27	27	32	2	8	5	3	1	4	2	4	4
28	28	30	4	5	1	1	1	1	2	2	4
29	29	26	2	2	5	1	2	4	2	5	5

Figure 8

SPSS Data View showing the cleaned and coded dataset used for analysis.'

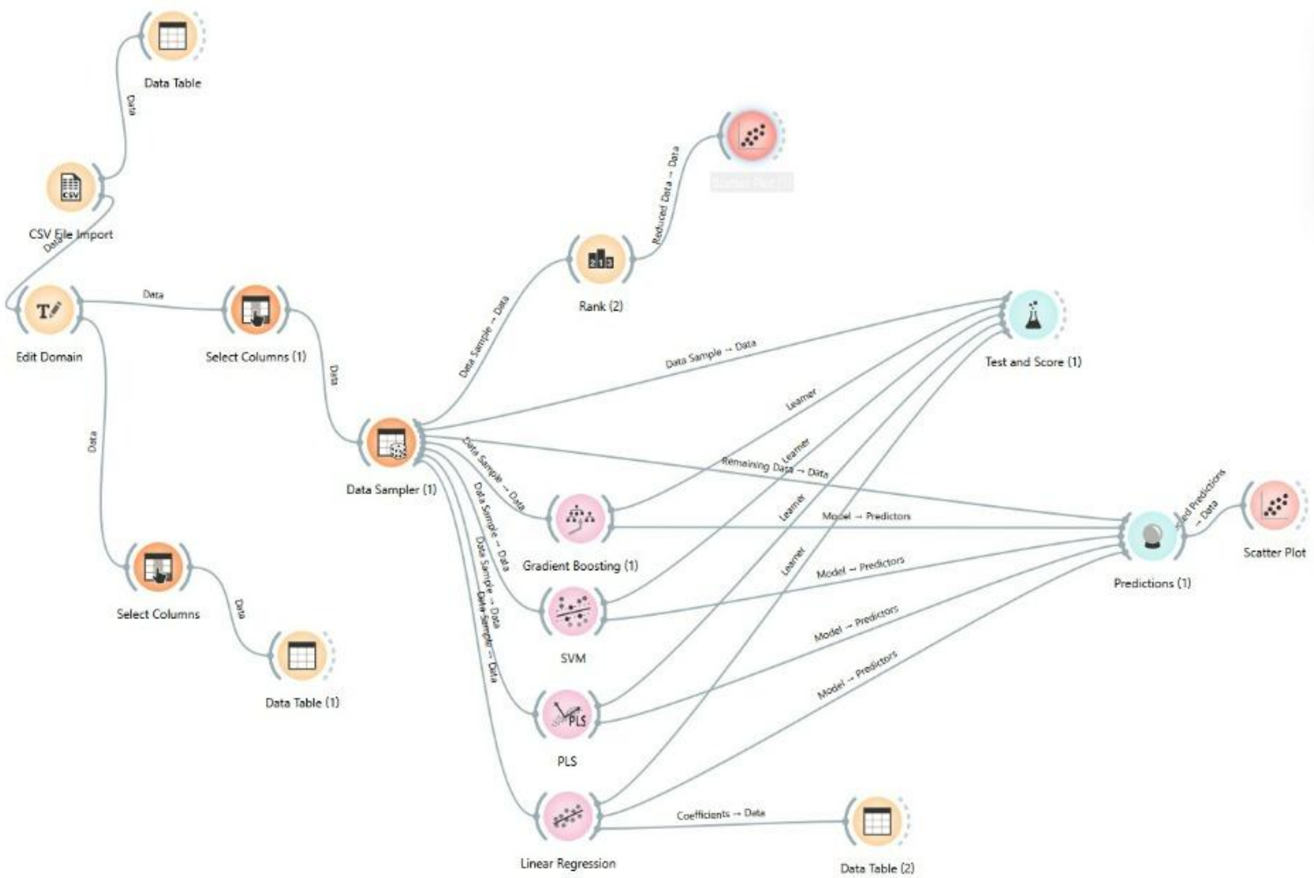


Figure 9 Machine Learning Workflow. This workflow shows how the dataset was processed and how SVM, Random Forest, Gradient Boosting, PLS, and Linear Regression models were applied and evaluated in Orange.

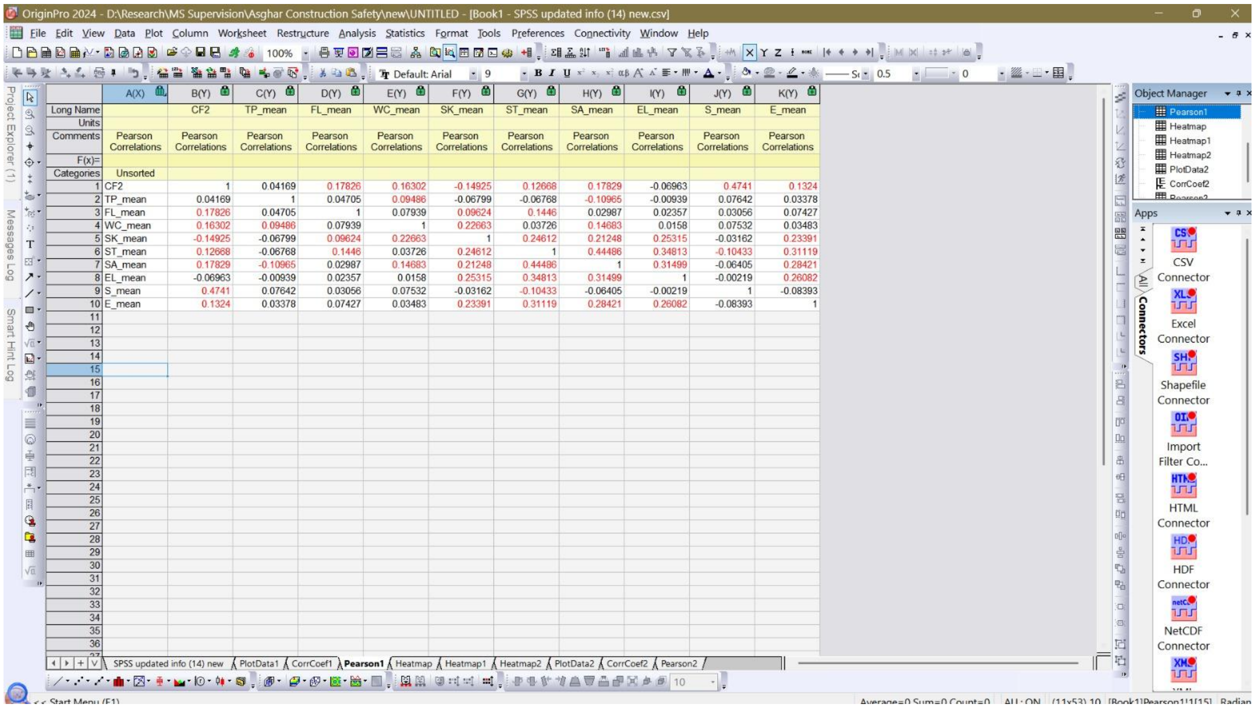


Figure 10. Pearson Correlation Matrix (OriginPro).

This screenshot shows the Pearson correlation values computed among all key variables used in the study.

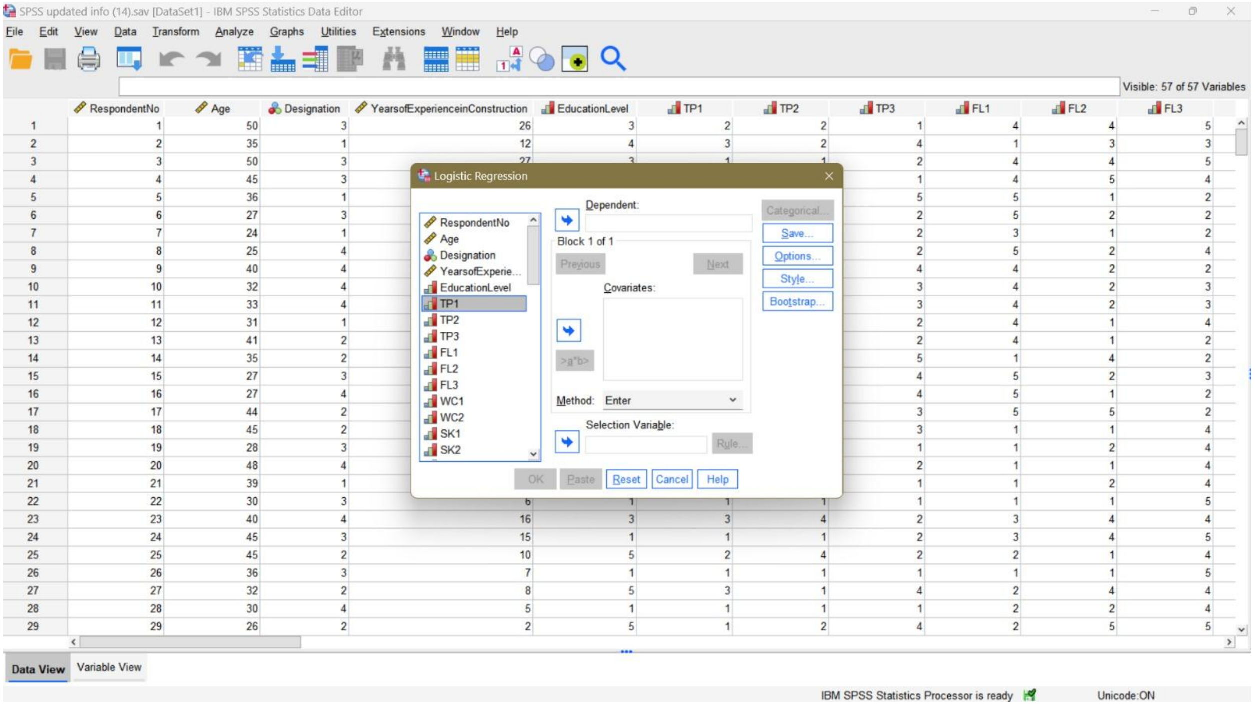


Figure 11. Logistic Regression Setup (SPSS Software)

This screenshot shows the SPSS interface used to configure and run the binary logistic regression model.

Settings

Sampling type: Stratified 20-fold Cross validation
Target class: None, show average over classes

Scores

Model	AUC	CA	F1	Prec	Recall	MCC
SVM (2)	0.863	0.757	0.756	0.760	0.757	0.516
Random Forest (1)	0.833	0.757	0.757	0.758	0.757	0.515
Gradient Boosting	0.839	0.757	0.757	0.758	0.757	0.514

Figure 12. ML Model Performance (20-Fold Cross-Validation)

Test and Score **Sun May 11 25, 13:55:59**

Settings

Sampling type: No sampling, test on training data
Target class: None, show average over classes

Scores

Model	AUC	CA	F1	Prec	Recall	MCC
SVM (2)	0.903	0.806	0.805	0.809	0.806	0.614
Random Forest (1)	0.989	0.951	0.951	0.952	0.951	0.903
Gradient Boosting	0.931	0.866	0.866	0.867	0.866	0.732

Figure 13. ML Model Performance (Training Data)

Test and Score Sun May 11 25, 13:56:40

Settings

Sampling type: No sampling, test on testing data
Target class: None, show average over classes

Scores

Model	AUC	CA	F1	Prec	Recall	MCC
SVM (2)	0.896	0.813	0.813	0.818	0.813	0.631
Random Forest (1)	0.881	0.780	0.780	0.783	0.780	0.563
Gradient Boosting	0.909	0.807	0.806	0.814	0.807	0.622

Figure 14 ML Model Performance (Testing Data Evaluation)

.