

RISK PERCEPTION MODELING BASED ON
PHYSIOLOGICAL AND EMOTIONAL RESPONSES

DING HUIZHE

FACULTY OF ENGINEERING
UNIVERSITI MALAYA
KUALA LUMPUR

2024

**RISK PERCEPTION MODELING BASED ON
PHYSIOLOGICAL AND EMOTIONAL RESPONSES**

DING HUIZHE

**THESIS SUBMITTED IN FULFILMENT OF THE
REQUIREMENTS FOR THE DEGREE OF DOCTOR OF
PHILOSOPHY**

**FACULTY OF ENGINEERING
UNIVERSITI MALAYA
KUALA LUMPUR**

2024

UNIVERSITI MALAYA
ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: DING HUIZHE

Matric No: 17021906/1

Name of Degree: DOCTOR OF PHILOSOPHY

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

RISK PERCEPTION MODELING BASED ON PHYSIOLOGICAL AND
EMOTIONAL RESPONSES

Field of Study:

RISK PERCEPTION, PHYSIOLOGICAL RESPONSES, EMOTIONAL
RESPONSES

I do solemnly and sincerely declare that.

- (1) I am the sole author/writer of this Work;
- (2) This Work is original;
- (3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
- (4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
- (5) I hereby assign all and every rights in the copyright to this Work to the Universiti Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
- (6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature

Date: 16/8/2024

Subscribed and solemnly declared before,

Witness's Signature

Date: 16/8/2024

Name:

Designation:

RISK PERCEPTION MODELING BASED ON PHYSIOLOGICAL AND EMOTIONAL RESPONSES

ABSTRACT

Risk perception refers to how individuals perceive objective risks. Although it initially emerged in social sciences, it has become a crucial aspect of safety science due to its significance in understanding unsafe behaviors. It can help safety managers develop a comprehensive understanding of risk based on traditional engineering risk assessment principles, facilitating the transition from the Safety I to Safety II paradigm. Therefore, accurate risk perception has become vital. This research aims to develop models for objectively assessing perceived risk. Previous studies have employed machine learning techniques to classify high and low-risk situations based on physiological responses. However, the performance of these algorithms in situations with closely comparable risk magnitudes remains uncertain. This issue is crucial as it directly impacts their practicality and generalization. To address this concern, four driving clips were selected as stimuli, including relatively low (1.87) and high (3.97) risk levels, as well as two clips with slight variations in their degree of riskiness (2.45 and 2.85, respectively). Fifty-five subjects were recruited to synchronously measure their physiological signals, including Electrodermal Activity (EDA), Heart Rate Variability (HRV), Pupil Diameter (PD), and Skin Temperature (ST). A Pleasure-Arousal-Dominance (PAD) model was used to induce and express mixed emotions. Subsequently, statistical analyses were performed to identify indicators that showed significant differences. These results varied significantly, including three emotional dimensions, two skin conductance indicators (EDR and EDL), and several ECG indicators (such as HF, LF/HF and A++) reflecting short-time changes. As the perceived risk level increased, subjects' emotions experienced more negative, arousal and a diminished sense of control. In terms of physiological changes, there was an increase in sympathetic activity and a concurrent decline in the

vagus nerve at a macro-level. However, the changes that resulted from consecutive heartbeats were characterized by rapid and erratic variations at a micro-level. Additionally, these observed significant differences were primarily attributed to variations in risk levels, rather than personal differences. In terms of feature importance, physiological and emotional indicators that showed significant differences or greater fluctuations demonstrated greater sensitivity. Finally, three base models, Artificial Neural Network (ANN), Random Forest (RF), Support Vector Classification (SVC), and two integrated models were trained to classify perceived risk using higher sensitivity features. The ANN demonstrated superior ability in distinguishing low and high-risk levels. However, when risk degrees were closely matched, the integrated model with weight adjustments based on base models outperformed ANN. To validate the research findings, a second experiment was conducted in a construction scenario, still utilizing two clips with closely matched risk degrees. It was demonstrated that the primary results derived from statistical analysis and machine learning modelling were remarkably consistent, thereby confirming the effectiveness and generalization of the proposed weight adjustment algorithm, particularly in situations with closely matched risk levels.

Keywords: Risk Perception, Physiological Responses, PAD Model, Statistical Analysis, Machine Learning.

PEMODELAN PERSEPSI RISIKO BERDASARKAN TINDAK BALAS

FISIOLOGI DAN EMOSI

ABSTRAK

Persepsi risiko merujuk kepada bagaimana individu mengenal pasti risiko objektif. Walaupun ia pada mulanya muncul dalam sains sosial, ia telah menjadi aspek penting dalam sains keselamatan kerana kepentingannya dalam memahami perilaku tidak selamat. Ia boleh membantu pengurus keselamatan membangunkan pemahaman menyeluruh tentang risiko berdasarkan prinsip penilaian risiko kejuruteraan tradisional, memudahkan peralihan dari paradigma Keselamatan I kepada Keselamatan II. Persepsi risiko yang tepat telah menjadi penting. Penyelidikan ini bertujuan untuk membangunkan model untuk menilai risiko yang dirasakan secara objektif. Kajian terdahulu telah menggunakan teknik pembelajaran mesin untuk mengelasifikasi situasi risiko tinggi dan rendah berdasarkan respons fisiologi. Walau bagaimanapun, tidak pasti sama ada algoritma ini berfungsi dengan baik dalam kes di mana magnitud risiko adalah relatif hampir. Isu ini adalah penting kerana ia memberi kesan secara langsung kepada praktikaliti dan amalan umum. Untuk mengatasi kebimbangan ini, empat klip pemanduan telah dipilih sebagai stimuli, termasuk tahap risiko yang rendah (1.87) dan tinggi (3.97), serta dua klip dengan variasi kecil dalam darjah risiko mereka (masing-masing 2.45 dan 2.85). Lima puluh lima subjek telah diambil untuk mengukur secara serentak isyarat fisiologi mereka, termasuk Aktiviti Elektrodermal (EDA), Variabiliti Denyut Jantung (HRV), Diameter Pupil (PD), dan Suhu Kulit (ST). Perasaan bercampur telah diinduksi dan dinyatakan dengan menggunakan model Kesenangan-Pengegaran-Ketuan (PAD). Seterusnya, analisis statistik telah dijalankan untuk mengenal pasti petunjuk yang menunjukkan perbezaan yang signifikan. Keputusan-keputusan ini berbeza secara ketara, termasuk tiga dimensi emosi, dua petunjuk pengaliran kulit (EDR dan EDL), dan beberapa petunjuk EKG (seperti HF, LF/HF, dan A++) yang mencerminkan perubahan dalam jangka masa pendek.

Apabila tahap risiko yang dirasakan meningkat, emosi subjek mengalami lebih banyak negatif, rangsangan, dan penurunan rasa kawalan. Dari segi perubahan fisiologi, terdapat peningkatan dalam aktiviti simpatetik dan penurunan serentak dalam saraf vagus pada tahap makro. Walau bagaimanapun, perubahan yang dihasilkan daripada degupan jantung berturut-turut dicirikan oleh variasi pantas dan tidak menentu pada tahap mikro. Selain itu, perbezaan ketara yang diperhatikan ini terutama disebabkan oleh variasi dalam tahap risiko, bukannya perbezaan peribadi. Dari segi kepentingan ciri, penunjuk fisiologi dan emosi yang menunjukkan perbezaan ketara atau turun naik yang lebih besar menunjukkan sensitiviti yang lebih tinggi. Akhirnya, tiga model asas, Rangkaian Neural Buatan (ANN), Hutan Rawak (RF), Pengelasan Vektor Sokongan (SVC), dan dua model bersepadu telah dilatih untuk mengklasifikasikan risiko yang dirasakan menggunakan ciri-ciri dengan sensitiviti yang lebih tinggi. ANN menunjukkan keupayaan unggul dalam membezakan tahap risiko rendah dan tinggi. Walau bagaimanapun, apabila tahap risiko hampir sama, model bersepadu dengan pelarasan berat berdasarkan model asas mengatasi prestasi ANN. Untuk mengesahkan penemuan penyelidikan, eksperimen kedua dijalankan dalam senario pembinaan, masih menggunakan dua klip dengan tahap risiko yang hampir sama. Ia menunjukkan bahawa hasil utama yang diperolehi daripada analisis statistik dan pemodelan pembelajaran mesin adalah sangat konsisten, sekali gus mengesahkan keberkesanan dan generalisasi algoritma pelarasan berat yang dicadangkan, terutamanya dalam situasi dengan tahap risiko yang hampir sama.

Kata Kunci: Persepsi Risiko, Tindak Balas Fisiologi, Model PAD, Analisis Statistik, Pembelajaran Mesin.

ACKNOWLEDGEMENTS

The journey to complete this thesis has undoubtedly been a formidable one, particularly as I navigated the intricate landscape of academic pursuits while embracing the joys and responsibilities of motherhood. The challenges inherent in this dual role were further compounded by the extraordinary circumstances of a global pandemic spanning three years. However, it is precisely these challenges that have enriched my perspective and fortified my resilience.

I extend my heartfelt gratitude to my supervisor, Dr. Raja Ariffin bin Raja Ghazilla and Prof. Ir. Ramesh Singh a/l Kuldip Singh for their unwavering support, guidance, and mentorship throughout my doctoral journey. Their profound knowledge, insightful feedback and constant encouragement have played a pivotal role in shaping the trajectory of my research and academic development. I consider myself truly fortunate to have had such a dedicated and inspiring mentor.

My deep thanks go to my family for their enduring love, understanding and encouragement. Their unwavering support has served as the pillar of strength that sustained me during the challenging phases of my doctoral studies. I am profoundly grateful for their sacrifices and unwavering belief in my abilities.

I am also indebted to my friends and colleagues who provided invaluable assistance, stimulating discussions, and a supportive academic community. Their camaraderie made the research process more enriching and enjoyable.

Thank you all for being an integral part of my academic and personal growth.

TABLE OF CONTENTS

Abstract	III
Abstrak	V
Acknowledgements	VII
Table of Contents	VIII
List of Figures	XV
List of Tables.....	XVIII
List of Symbols And Abbreviations.....	XX
List of Appendices	XXIII
CHAPTER 1: INTRODUCTION.....	1
1.1 Background.....	1
1.1.1 The Emergence of Risk Perception in Social Science.....	1
1.1.2 The Development of Risk Perception in Safety Science.....	2
1.1.3 The Assessment of Perceived Risk.....	3
1.2 Problem statement	5
1.3 Research Objective	5
1.4 Research Scope	6
1.5 Thesis Structure	6
CHAPTER 2: LITERATURE REVIEW.....	8
2.1 Introduction.....	8
2.2 Risk Perception.....	8
2.2.1 The definition of Risk Perception.....	8
2.2.2 The Process of Perceiving Risk.....	9
2.3 Assessment of Perceived Risk	11

2.3.1	Risk Perception in Social Science	11
2.3.1.1	The existences and determinants of risk perception.....	11
2.3.1.2	The assessment paradigm of risk perception.....	13
2.3.2	Risk Perception in Safety Science.....	14
2.3.2.1	The existence of risk perception in safety science	14
2.3.2.2	The biases and influence factors of risk perception in safety science	15
2.3.2.3	The significance of risk perception in safety management	17
2.3.2.4	The assessment of perceived risk in safety science.....	18
2.3.3	The Significant Role of Risk Perception in Emergency.....	22
2.3.4	Risk Perception in Ergonomics	24
2.3.4.1	Cognitive assessment and its comparison with risk perception	25
2.3.4.2	Human reliability assessment and its comparison with risk perception	26
2.3.5	Comparison of Different Risk Perception Assessment Paradigms	28
2.4	Physiological and Emotional Responses in Risk Perception.....	30
2.4.1	Electrodermal Responses.....	31
2.4.1.1	The commonly used indicators of EDA	31
2.4.1.2	The changes of EDA during risk perception	31
2.4.2	Cardiac Responses.....	32
2.4.2.1	The commonly used indicators of HRV.....	33
2.4.2.2	The changes of HRV during risk perception.....	33
2.4.3	Eye Tracking Behaviors	34
2.4.3.1	The basic eye-tracking indicators.....	34
2.4.3.2	The application of eye-tracking during risk perception	35
2.4.4	Skin Temperature Changes.....	36

2.4.5	Emotional Responses.....	40
2.4.5.1	The induced complicated emotions during risk perception.....	41
2.4.5.2	The expression of induced emotions	41
2.4.6	Discussion of Physiological and Emotional Responses	42
2.5	Application of Machine Learning.....	43
2.5.1	Application of Machine Learning on Risk Assessment	43
2.5.1.1	Application phase for ML	43
2.5.1.2	Data sources for ML.....	45
2.5.1.3	The frequency of method usage	46
2.5.2	Application of Machine Learning on Human Factors.....	46
2.5.3	Application of Machine Learning on Perceived Risk	51
2.5.3.1	The application of ML for assessing perceived risk in social science	52
2.5.3.2	The application of ML for assessing perceived safety risk	52
2.5.4	Discussion of Machine Learning Application.....	55
2.5.5	The Comparison of ML with Other Methods.....	56
2.6	Individual Differences of Risk Perception.....	57
2.7	Gap Summary	59
CHAPTER 3: METHODOLOGIES		61
3.1	Introduction.....	61
3.2	The Comprehensive Research Design	61
3.3	Stimulus	62
3.4	Participants	65
3.5	Measurements	65
3.5.1	Physiological Measurements	65
3.5.2	Emotion Measurements	66

3.6	Procedure	68
3.7	Apparatus	69
3.7.1	Software of ErgoLAB.....	70
3.7.2	Hardware of ErgoLAB	71
3.7.3	Physiological indices	74
3.8	Validation Experiment.....	75
3.9	Data Analysis.....	77
3.9.1	Statistical Analysis	77
3.9.1.1	Discrete analysis of PAD emotions.....	77
3.9.1.2	Continuous analysis of physiological responses	79
3.9.1.3	Personal difference	81
3.9.2	Machine Learning.....	82
3.9.2.1	Scaled raw dataset	83
3.9.2.2	Collinearity detection and feature importance	84
3.9.2.3	Split Dataset	85
3.9.2.4	Model selection	86
3.9.2.5	Classification.....	86
3.9.2.6	Integration of trained models	88
3.10	Summary.....	90

CHAPTER 4: STATISTICAL ANALYSIS OF PHYSIOLOGICAL AND EMOTIONAL RESPONSES	92	
4.1	Introduction.....	92
4.2	Statistical Analysis of PAD	92
4.3	Statistical Analysis of Physiological Data.....	94
4.3.1	Statistical analysis of EDA.....	94
4.3.2	Statistical analysis of HRV.....	95

4.3.3	Statistical of PD and ST	100
4.4	Discussion of Emotional and Physiological Responses	101
4.4.1	Discussion of Emotional Responses.....	101
4.4.1.1	Emotional state changes	101
4.4.1.2	Differentiation ability among three dimensions.....	102
4.4.2	Discussion of Physiological Responses.....	102
4.4.2.1	The changes of EDA	102
4.4.2.2	The changes of ECG	104
4.4.2.3	The changes of PD and ST	110
4.5	Personal Differences	111
4.5.1	Personal difference of PAD.....	111
4.5.2	Personal difference of EDA.....	112
4.5.3	Personal difference of HRV	113
4.5.4	Discussion of Personal Difference	113
4.6	Summary.....	114

CHAPTER 5: MACHINE LEARNING FOR PERCEIVED RISK ASSESSMENT

.....	117	
5.1	Introduction.....	117
5.2	Feature Importance	117
5.2.1	Correlation of input variables.....	117
5.2.2	Recursive Feature Elimination (RFE) to select variables.....	118
5.2.3	Calculation of Feature importance	119
5.3	Perceived Risk Classification	120
5.4	Results of Machine Learning.....	123
5.4.1	Fitting Evaluation	123
5.4.1.1	Fitting performance of RF	123

5.4.1.2	Fitting performance of SVC	124
5.4.1.3	Fitting performance of ANN	126
5.4.2	Based Model Performance	127
5.4.3	Integrated Models Performance	129
5.5	Discussion.....	131
5.6	Summary.....	133

**CHAPTER 6: VALIDATION FOR GENERALIZED CAPABILITY OF
MACHINE LEARNING 135**

6.1	Introduction.....	135
6.2	The Application Paradigm of Risk Perception Assessment	135
6.3	Validation for Statistical Analysis Results	138
6.3.1	Validating PAD Analysis Results	138
6.3.2	Validating Physiological Analysis Results.....	140
6.3.2.1	Statistical analysis of EDA.....	140
6.3.2.2	Statistical analysis of HRV	141
6.3.2.3	Statistical of PD and ST	145
6.3.3	Discussion and Comparison with Previous Statistical Findings	146
6.3.3.1	Discussion and comparison with emotional responses	146
6.3.3.2	Discussion and comparison with physiological responses.....	147
6.3.4	Confirmation of Personal Difference Results.....	148
6.3.4.1	Personal difference of PAD.....	148
6.3.4.2	Personal difference of EDA	148
6.3.4.3	Personal difference of HRV	149
6.3.4.4	Discussion and Comparison with Personal Difference	149
6.4	Validation For Machine Learning Performance	150
6.4.1	Comparison of Feature Importance	150

6.4.1.1	Correlation of input variables.....	150
6.4.1.2	RFE to select variables.....	151
6.4.1.3	Feature importance.....	151
6.4.2	Fitting Evaluation.....	152
6.4.3	Comparison of Based Model Performance.....	156
6.4.4	Comparison Integrated Models Performance.....	158
6.5	Comparison And Discussion With Previous Machine Learning.....	159
6.6	Summary.....	160
 CHAPTER 7: CONCLUSION AND RECOMMENDATION		163
7.1	Conclusions.....	163
7.1.1	Characterization of physiological and emotional changes in risky situations.....	163
7.1.2	Machine learning model to assess subjective perceived risk	165
7.1.3	Validate the performance of the proposed machine learning model.....	165
7.2	Significance of Study.....	166
7.2.1	Significance of Knowledge	166
7.2.2	Significance of Application.....	166
7.3	Limitations.....	167
7.4	Recommendations for Future Work	168
	References.....	170
	List of Publications and Papers Presented	219
	Appendix.....	220

LIST OF FIGURES

Figure 2.1: Conceptual safety management process	17
Figure 2.2: The process of how humans deal with emergency situations.....	23
Figure 2.3: The number of papers concerning the three risk assessment phases	44
Figure 2.4: The three types of data sources to build the model	45
Figure 2.5: Ten frequently used machine learning methods in risk assessment	46
Figure 2.6: Formulation of research problems based on the literature review.....	60
Figure 3.1: The comprehensive research flow chart	62
Figure 3.2: The Self-Assessment Manikin (SAM) visuals	67
Figure 3.3: Diagram of experimental procedure	69
Figure 3.4: The image of computer host and LED monitor	72
Figure 3.5: The different physiological measurement sensors.....	72
Figure 3.6: The wireless ECG sensor and its wearing methods.....	73
Figure 3.7: The wireless EDA sensor and its wearing methods	73
Figure 3.8: The Screen-based eye-tracking device	73
Figure 3.9: The wireless ST sensor and its wearing methods	74
Figure 3.10: The statistical analysis process	80
Figure 3.11: The machine learning process for classifying perceived risk levels and determining feature importance	83
Figure 3.12: The integration of the computational process of trained models through weights calculation for different categories	89
Figure 3.13: The integration of the computational process of trained models through the stacking technique	90
Figure 4.1: The changing trend of P, A, and D with risk ranks in driving.....	93
Figure 4.2: Mean values of HRV frequency-domain indicators with risk level	98
Figure 4.3: Mean values of HRV nonlinear indicators with risk level	100

Figure 5.1: The heatmap of twenty input features in driving.....	118
Figure 5.2: Feature selection results of RFE cross-validation in driving.....	119
Figure 5.3: Feature importance of different indicators in driving using Gini.....	119
Figure 5.4: Feature importance of different indicators in driving using permutation...	120
Figure 5.5: The implementation process for three basic machine learning models.....	122
Figure 5.6: The learning curve of RF with a 0.7:0.15:0.15 ratio split for training, validation, and test datasets in driving.....	123
Figure 5.7: The learning curve of RF with a 0.5:0.25:0.25 ratio split for training, validation, and test datasets in driving.....	124
Figure 5.8: The learning curve of SVC for driving experiment data with a 0.7:0.15:0.15 ratio split for training, validation, and test datasets in driving.....	125
Figure 5.9: The learning curve of SVC for driving experiment data with a 0.5:0.25:0.25 ratio split for training, validation, and test datasets in driving.....	125
Figure 5.10: The learning curve of ANN for driving experiment data with a 0.7:0.15:0.15 ratio split for training, validation, and test datasets in driving.....	126
Figure 5.11: The learning curve of ANN for driving experiment data with a 0.5:0.25:0.25 ratio split for training, validation, and test datasets in driving.....	126
Figure 5.12: Confusion matrix and result reports of three models with two ratios split for training, validation, and test datasets in driving.....	128
Figure 5.13: The ROC curves and AUC values of three models with two ratios split for training, validation, and test datasets in driving.....	129
Figure 5.14: Confusion matrix and result reports of integrated model.....	130
Figure 5.15: The ROC curves and AUC values of the integrated model with stacking technique.....	131
Figure 6.1: The application paradigm for assessing perceived risk through physiological and emotional responses.....	136
Figure 6.2: The changing trend of P, A, and D with risk ranks in construction.....	139
Figure 6.3: Mean values of HRV frequency-domain indicators with risk level in construction.....	144
Figure 6.4: The heatmap of twenty input features in construction.....	150

Figure 6.5: Feature selection results of RFE cross-validation in construction	151
Figure 6.6: Feature importance of different indicators in construction using Gini.....	152
Figure 6.7: Feature importance of different indicators in constructing using permutation	152
Figure 6.8: The learning curve of RF with a 0.7:0.15:0.15 ratio split for training, validation, and test datasets in construction.....	153
Figure 6.9: The learning curve of RF with a 0.5:0.25:0.25 ratio split for training, validation, and test datasets in construction.....	153
Figure 6.10: The learning curve of SVC with a 0.7:0.15:0.15 ratio split for training, validation, and test datasets in construction.....	154
Figure 6.11: The learning curve of SVC with a 0.5:0.25:0.25 ratio split for training, validation, and test datasets in construction.....	154
Figure 6.12: The learning curve of ANN with a 0.7:0.15:0.15 ratio split for training, validation, and test datasets in construction.....	154
Figure 6.13: The learning curve of ANN with a 0.5:0.25:0.25 ratio split for training, validation, and test datasets in construction.....	155
Figure 6.14: Confusion matrix and result reports of three models with two ratios split for training, validation, and test datasets in construction	156
Figure 6.15: The ROC curves and AUC values of three models with two ratios split for training, validation, and test datasets in construction	157
Figure 6.16: Confusion matrix and result reports of integrated model in constructing	158
Figure 6.17: The ROC curves and AUC values of the integrated model with stacking technique in construction	159

LIST OF TABLES

Table 2.1: The common physiological indicators and their meanings.....	36
Table 2.2: ML application for different phases of risk assessment.....	44
Table 2.3: Description of published papers using machine learning to analyze human factors.....	47
Table 2.4: The comparison of ML with other methods used in risk perception	57
Table 3.1: A simple description of four driving clips	64
Table 3.2: Parameters of EDA signal preprocessing methods.....	70
Table 3.3: Parameters of ECG signal preprocessing methods	71
Table 3.4: Parameters of eye tracking raw data preprocessing methods	71
Table 3.5: A simple description of four construction clips.....	76
Table 4.1: Non-parameter test and pair comparison results of P, A, and D in driving...	92
Table 4.2: Description and repeated-measure ANOVA analysis results of EDA in driving	95
Table 4.3: Description and repeated-measure ANOVA analysis results of HRV time-domain indicators in driving	96
Table 4.4: Description and repeated-measure ANOVA analysis results of HRV frequency-domain indicators in driving	97
Table 4.5: Description and repeated-measure ANOVA analysis results of HRV nonlinear indicators in driving	98
Table 4.6: Description and repeated-measure ANOVA analysis results of PD and ST in driving	101
Table 4.7: Result of hierarchical linear modeling for PAD in driving.....	111
Table 4.8: Result of hierarchical linear modeling for EDR in driving.....	112
Table 4.9: Result of hierarchical linear modeling for HRV in driving	113
Table 4.10: The main outcomes and limitations obtained from statistical analysis.....	116

Table 5.1: The results of grouping the feature importance based on Gini and Permutation methods in driving.....	120
Table 5.2: The optimized parameters of RF.....	124
Table 5.3: The optimized parameters of SVC.....	125
Table 5.4: The optimized parameters of ANN.....	127
Table 5.5: The main outcomes and limitations obtained from modeling	133
Table 6.1: Non-parameter test and pair comparison results of P, A, and D in construction	138
Table 6.2: Description and repeated-measure ANOVA results of EDA in construction	141
Table 6.3: Description and repeated-measure ANOVA results of HRV time-domain indicators in construction	142
Table 6.4: Description and repeated-measure ANOVA results of HRV frequency-domain indicators in construction	143
Table 6.5: Description and repeated-measure ANOVA analysis results of HRV nonlinear indicators in construction	145
Table 6.6: Description and repeated-measure ANOVA analysis results of PD and ST in construction	146
Table 6.7: Result of hierarchical linear modeling for PAD in construction	148
Table 6.8: Result of hierarchical linear modeling for EDR in construction	149
Table 6.9: Result of hierarchical linear modeling for HRV in construction.....	149
Table 6.10: The optimized parameters of RF in construction.....	155
Table 6.11: The optimized parameters of SVC in construction.....	155
Table 6.12: The optimized parameters of ANN in construction.....	155
Table 6.13: The main outcomes and limitations obtained from comparing the first and validated experimental analysis results.....	161

LIST OF SYMBOLS AND ABBREVIATIONS

ANN	:	Artificial Neural Network
ANOVA	:	Analysis of Variance
ANS	:	Autonomic Nervous System
AUC	:	Area Under the Curve
bpm	:	Beats Per Minute
BT	:	Bagging Tree
CREAM	:	Cognitive Reliability and Error Analysis Method
DT	:	Decision Tree
ECG	:	electrocardiogram
EDA	:	Electrodermal Activity
EDL	:	Electrodermal Level
EDR	:	Electrodermal Response
EEG	:	Electroencephalogram
FAHP	:	Fuzzy Analytic Hierarchy Process
FFD	:	Fitness-for-duty
FMEA	:	Failure Mode and Effects Analysis
FTA	:	Fault tree analysis
GSVM	:	Gaussian Support Vector Machine
HAZOP	:	Hazard and operability analysis
HCA	:	Human cognitive assessment
HEA	:	Human error analysis
HF	:	high frequency
HLM	:	Hierarchical Linear Modeling
HPA	:	Human performance assessment

HR	:	Heart Rate
HRA	:	Human reliability Analysis
HRA	:	Human Reliability Assessment
HRV	:	Heart Rate Variability
IB	:	Inattentive Blindness
IBI	:	Inter-Beat Interval
KNN	:	K-Nearest Neighbor
LED	:	Light Emitting Diode
LF	:	low frequency
LR	:	Logistic Regression
LSTM	:	Long Short-Term Memory-Based Method
MDI	:	Mean Decrease Impurity
ML	:	Machine Learning
MLP	:	Multilayer Neural Networks
ms	:	Milliseconds
ms^2	:	Square Millisecond
NLP	:	Natural Language Processing
NN	:	Neural Networks
NPPs	:	Nuclear Power Plants
PAD	:	Pleasure-Arousal-Dominance
PD	:	Pupil Size
PI	:	Permutation Importance
PNS	:	Parasympathetic Nervous Systems
PPE	:	Personal protective equipment
PSFs	:	Performance Shaping Factors
PPG	:	Photoplethysmogram

RF	:	Random Forest
ROC	:	Receiver Operating Characteristic
SAM	:	Self-Assessment Manikin
SI	:	Safety Intelligence
SA	:	Situation Awareness
SCL	:	Skin Conductivity Levels
SLIM	:	Success Likelihood Index Method
SMS	:	Safety management system
SNS	:	Sympathetic Nervous System
SPAR-H	:	Standardized Plant Analysis Risk Human reliability analysis
ST	:	Skin Temperature
SVC	:	Support Vector Classification
SVM	:	Support Vector Machine
THERP	:	Technique for Human Error Rate Prediction
ULF	:	Ultra-low-frequency
VLF	:	Very-low-frequency
VR	:	Virtual Reality

LIST OF APPENDICES

Appendix A: Photos of the driving and construction scenarios experiments	220
Appendix B: Links to the raw data of the two experiments.....	221
Appendix C: The Confusion matrix and result reports of RF with two ratios split for training and validation datasets in driving and construction.....	222
Appendix D: The ROC curves and AUC values of RF with two ratios split for training and validation datasets in driving and construction	223
Appendix E: The Confusion matrix and result reports of SVC with two ratios split for training and validation datasets in driving and construction.....	224
Appendix F: The ROC curves and AUC values of SVC with two ratios split for training and validation in driving and construction.....	225
Appendix G: The Confusion matrix and result reports of ANN with two ratios split for training and validation datasets in driving and construction.....	226
Appendix H: The ROC curves and AUC values of ANN with two ratios split for training and validation datasets in driving and construction	227

CHAPTER 1: INTRODUCTION

1.1 Background

Human factors have been a central focus of research in various industries since the late 1980s. They are recognized as the primary cause of accidents (Bellamy et al., 2008). Notably, statistical analysis of accidents has shown that unsafe human behaviors contribute to at least 66% and, in some cases, over 90% of incidents across various industries (Azadeh & Zarrin, 2016). The concept of unsafe human behaviors emerged from the interaction between workers' decision-making processes and hazardous work conditions (Wang et al., 2016). However, preceding unsafe human behaviors and decisions is the individual's perception of risk (Chan et al., 2022; Hill et al., 1997; Sitkin & Weingart, 1995; Tixier et al., 2014; Weber et al., 2002). The viewpoint that defects in risk perception were recognized as crucial cause of accidents dates back to as early as 1992 (Rundmo, 1992b).

1.1.1 The Emergence of Risk Perception in Social Science

The concept of risk perception emerged in the 1960s within the policy domain (Sjöberg et al., 2004). It refers to individuals' recognition and understanding of various objective risks within their environment (Botterill & Mazur, 2004; Slovic, 1988; Slovic et al., 1982; Veland & Aven, 2013). A significant gap was found between risk assessments made by experts and those held by the public. For example, experts may consider living near a nuclear facility to be less risky than ordinary activities such as drinking or driving, but society does not share this perspective. This inconsistency has encouraged prompt exploration in social science and psychology to explore the underlying reasons (Wilson et al., 2019). In recent years, divergent public perceptions of the same risk have continued to exist in the context of both the COVID-19 pandemic (Dryhurst et al., 2022) and the Fukushima nuclear leakage incident (Morioka, 2014). This disparity significantly

influences the development of policies and strategies for risk communication in risk management (Loewenstein et al., 2001; Sjöberg et al., 2004; Slovic et al., 2004).

Additionally, extensive research has been conducted on influencing factors. For example, gender, age, prior knowledge, and life experiences can all impact an individual's subjective risk perception (Shou & Olney, 2021; Wachinger et al., 2013). In particular, George F. Loewenstein and his team (Loewenstein et al., 2001) delved deeper by introducing the "Risk as Feelings" hypothesis, which emphasized the role of emotions in risk perception.

1.1.2 The Development of Risk Perception in Safety Science

Risk perception was introduced and systematically studied, gaining significant recognition in safety science in the 1990s (Flin & Mearns, 1994; Okrent, 1998; Rundmo, 1992a, 1992b, 1996). A series of studies have provided supporting evidence of the impact of risk perception on decision-making and behavior across various domains, including construction (B. Choi et al., 2019; B. G. Lee et al., 2021), transportation (Harbeck & Glendon, 2018; Herrero-Fernandez et al., 2016; Sohail et al., 2023; Stülpnagel et al., 2022), aviation (Ji et al., 2011), natural disaster (Ng, 2023; Paton et al., 2000), emergency evacuation and preparedness (Choi et al., 2016; Riad et al., 1999). It implies that accurate risk perception is essential for making informed decisions and conducting appropriate behaviors (Rundmo, 1996; Taylor & Snyder, 2017).

In the industrial sector, biases in perceived risk can also arise. These differences may arise from disparities between beliefs about risks and the outcomes of traditional technological risk assessments (Aven, 2018; Aven & Kristensen, 2005; Aven & Renn, 2010), or they may result from differing risk perceptions among various personnel groups within the same organization, such as management and frontline operators (Hon et al., 2023; Mearns & Flin, 1995), and even among different employee groups (De Salvo et al., 2022; Gürcanlı et al., 2015; Hon et al., 2023). Any of these discrepancies can lead to

inaccurate decision-making (Loewenstein et al., 2001; Slovic et al., 2004) and behaviors (Taylor, 1964; Wilde, 2014), ultimately undermining system reliability (Zhang et al., 2022). For example, when individuals perceive potential harm as high, they are more inclined to be cautious and engage in proactive and appropriate behaviors to reduce risk. Conversely, if they perceive them as low, they may disregard the hazard and engage in unsafe actions (B. G. Lee et al., 2021).

Biases in perceiving risk can offer valuable insight for safety managers and professionals. The variations are considered significant from the current perspective (Brown, 2014). Incorrect or inappropriate risk perception from the individual level can provide safety managers with an entirely different perspective on risk information. Many safety measures aimed at risk reduction are predominantly based on the outcomes of expert or traditional engineering risk assessments. However, when the reduced risk level is higher than the perceived risk at the individual level, the risk may still be perceived as exceeding the individual's acceptable threshold from a personal perspective. So, understanding risk from an individual perceiving perspective can complement the existing risk assessment. Consequently, accurate assessment of individual risk perceptions becomes essential.

1.1.3 The Assessment of Perceived Risk

Numerous areas have developed methods to measure perceived risk. In social science, the psychometric paradigm was used to quantitatively assess risk in the 1980s and 1990s (Slovic, 1990). This assessment scale could identify and quantify the similarities and differences in risk cognition and attitudes among groups, but its effectiveness at the individual level was limited (Sjöberg et al., 2004).

Additionally, the inherent subjectivity sets it apart from traditional or engineering risk assessment in safety sciences. The former frequently utilize combinations of probability and consequences to qualitatively or quantitatively depict risk carried out by safety

experts or engineering (Aven & Kristensen, 2005; Marhavilas et al., 2011). However, it is vitally important to note that these methods often prioritize hazards to machinery and the environment, frequently overlooking potential risks to human factors (Mearns & Flin, 1995; Pasman et al., 2013). More specifically, the risk perception of individuals directly involved in risky situations is often disregarded (Aven, 2018; Siegrist & Árvai, 2020).

Risk perception was regarded as an influencing factor of human reliability in the human-machine system. A series of studies, including Human Reliability Analysis (HRA), Human Error Analysis (HEA), Human Performance Assessment (HPA), and Human Cognitive Assessment (HCA), have been conducted. HRA, proposed in the early 1980s, evaluates the contribution of operators to system safety performance (La Fata et al., 2021).

Both safety science and ergonomics assessment methods rely on accident historical data and expert experience in system operations, providing insights into objective risks (Chang et al., 2016; Liu et al., 2018). Therefore, it becomes evident that these two approaches are inadequate for quantifying risk perception. Moreover, both directions carried an underlying assumption that individuals who demonstrated reliability in standard work conditions would also perform effectively in risky situations. Nevertheless, it is essential to consider that a person's mental workload during emergencies can be twice as high (Connelly, 1997), compounded by the urgent and critical nature of risk scenarios (Woodcock & Au, 2013) and the presence of time pressure (Coeugnet et al., 2013). Consequently, perceiving risk was inevitably influenced, and the likelihood of human errors or unsafe behaviors was significantly increased in real-world scenarios.

In recent years, a new paradigm has emerged for assessing subjective risk perception, driven by advancements in psychophysiological measurement and machine learning (B. Choi et al., 2019; B. G. Lee et al., 2021). Initially, physiological response data associated with the risk perception process were obtained through measurements of physiological signals. Subsequently, these data were utilized as inputs to train machine learning models,

with the output representing the perceived level of risk. This approach brings about a novel opportunity to objectively evaluate perception-based risks that are inherently subjective and concealed.

1.2 Problem Statement

In industrial practices, risk assessment is the cornerstone of risk management systems (Manuele, 2021), guiding the development of safety decisions and strategies for risk mitigation. However, traditional engineering risk assessment methods predominantly present a particular challenge: the assessment conducted by professionals differs from the risk perceived by workers involved in risk situations (Arezes & Miguel, 2008). Moreover, perceived risk is highly subjective and concealed, challenging the traditional reliance on pure "numeric trust" (Krige, 1997). Hence, a scientific and objective assessment of their risk perception, complementing engineering risk assessment, is crucial for effective risk management (Manar et al., 2019; Namian et al., 2018) and the transition from Safety Management I to Safety Management II (Aven. & Terje., 2022).

1.3 Research Objective

The primary objectives of the research are to develop a suitable model to assess subjective perceived risk from the perspective of physiology and emotional responses, which corresponds closely to the risk level in humans. To achieve this goal, the following objectives have been established:

Objective 1: To characterize the physiological and emotional changes that arise in response to various risky situations, employing statistical methods for analysis.

Objective 2: To develop machine learning models to assess perceived risk levels based on physiological and emotional responses.

Objective 3: To validate the performance of the proposed machine learning model in assessing perceived risk levels.

1.4 Research Scope

The categorization of risk in risk perception varies between social science and safety science. In social science, the term 'risk' includes natural, environmental, infectious diseases and technological risks that extend beyond the scopes of production enterprises. Conversely, in safety science, the risk primarily refers to technological risks associated with the production process within an enterprise. This study specifically aims to investigate safety risks within the domain of safety science and focus on modeling perceived risk using machine learning technique based on physiological and emotional responses.

1.5 Thesis Structure

The thesis consists of seven chapters. The present chapter, **Chapter 1**, mainly introduces the research background. It provides an overview of risk perception and its importance in accident prevention. The discussion on research progress and limitations in risk perception leads to the formulation of the research objectives in this thesis.

Chapter 2 provides a literature review on risk perception, possessing an interdisciplinary nature, which covers sociology, safety science and ergonomics. Consequently, Chapter 2 introduces its concept, origins, primary research findings and limitations within these three domains. Given that this study employs physiological and emotional responses to represent risk perception, the following sections explore these responses in risky situations, potential individual differences and the application of machine learning in risk assessment. Finally, an overview of research gaps is provided.

Chapter 3 presents the complete experimental design and the subsequent data analysis methods, covering stimulus and participant screening, physiological and emotional measurement methods, experiment equipment, procedure, statistical analysis methods and the selection of machine learning algorithms. Additionally, this chapter also includes experimental validation.

Chapter 4 provides the statistical analysis of physiological and emotional data for the first experiment. It examines the differences in physiological and emotional responses induced at different risk levels. Subsequently, the source of significant differences is analyzed based on risk levels or individual factors. It aids in gaining initial insights into the patterns of response variations and the underlying neural regulatory mechanism across different risk situations.

Chapter 5 presents the process of detecting feature sensitivity and eliminating collinearity features, then using the new features as input vectors for classifying risk. Three basic models and two integrated models are developed. Finally, their performance is assessed and compared to select the optimal model.

Chapter 6 presents the application paradigm of risk perception assessment. Subsequently, experiment is conducted based on this framework in a new scenario. It validates the findings obtained from the initial study, including the statistical analysis results, feature importance and the optimal model. The aim is to confirm the generalization capacity of the optimal risk perception model in different scenarios.

Chapter 7 provides the conclusion and recommendations.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

While risk perception initially emerged in social science research, it has shown to be invaluable in the field of industry safety, as demonstrated by a well-known incident in the 2010 explosion and sinking of the Deepwater Horizon, resulting in the world's worst offshore oil spill (Reader & O'Connor, 2014). The investigation into this accident revealed that operational personnel failed to detect abnormalities in pressure readings and fluid quantities during negative pressure testing and oil well monitoring of oil and gas into the well (Bly, 2011). The primary reasons for these lapses in both job roles were their flawed risk perception, sometimes leading to unintentional disregard of hazards in complex and highly uncertain situations (Manar et al., 2019). This delayed and inaccurate risk perception represented a classical hazard that traditional risk assessment methods struggled to identify, particularly when applied in environments marked by deep uncertainty and complexity.

Given the importance of risk perception, achieving a comprehensive and profound understanding is crucial. Therefore, this chapter briefs on an overview of key research findings on risk perception from sociology, safety science, and ergonomics. Subsequently, building upon the established research findings, we seek to pinpoint areas where further research is needed. Moreover, exploration into the potential application of physiological and emotional measurements, along with machine learning algorithms, is undertaken in this study, offering relevant introductions.

2.2 Risk Perception

2.2.1 The definition of Risk Perception

Before delving into the concept of 'risk perception', it is essential to understand the term 'risk' itself. The concept of 'risk' emerged in academic literature as early as the 1980s (Marshall, 2020). Risk was typically defined as a combination of the probability or

frequency of a defined hazard occurring and the magnitude of the consequences resulting from the occurrence in the engineering industry (Aven & Kristensen, 2005; P.K. Marhavilas, 2011). However, the definition of risk in social science research differed from the engineering perspective. Social science recognizes that 'risk' is multidimensional and shaped by various factors, including personal and cultural influences, as well as the inherent characteristics of the risk itself (Pidgeon, 1998). For example, the risk of 'catastrophic potential' is more significant in the context of nuclear and chemical technology risks but less so in automobile travel (Gardner & Gould, 1989). Social science introduced the concept of 'risk perception' to express the multidimensional characteristics of risk. This term encapsulates the subjective understanding of risk held by humans in response to a wide range of hazards.

Risk perception generally refers to an individual's subjective judgment and evaluation of risks in situations involving danger or emergencies (Fung et al., 2010; Hallowell, 2010; Rundmo, 2000). According to Deery's study (Deery, 1999), the concept was mentioned in various theoretical models. For example, in an information-processing model proposed by Wickens, Hollands, Banbury, and Parasuraman (Wickens et al., 2015), risk perception, along with sensory processing, response selection, and response execution, played an important role in collecting relevant information and forming a perceived risk. Similarly, the Situation Awareness (SA) model (Endsley, 1995) emphasized the importance of perceiving elements in the environment as the first step, followed by developing a comprehensive understanding of the current situation and making predictions about future conditions.

2.2.2 The Process of Perceiving Risk

In the process of risk perception, Choi and Lee (2018) considered risk perception as the initial step, followed by risk assessment and safety behavior. Another perspective on risk perception and behavioral decision-making was provided by the dual-processing

theory (Chaiken & Trope, 1999), which suggests that the human brain operates in two modes when perceiving risk and making a decision: System 1 and System 2. The former is a fast, unconscious, instinctive, emotional and automatic mode of thinking that allows individuals to make conclusions with little effort (Aven, 2018). Conversely, System 2 entails slow, conscious, and deliberate thinking, demanding substantial mental effort (Janoff-Bulman, 2001; Smith & DeCoster, 2000). It is further supported by physiological experiments that have confirmed the existence of System 1 and System 2. Electroencephalogram (EEG) analysis identified two stages: the detection stage (marked by P200) and the evaluation stage (marked by LPP). The former was characterized by rapid, affective, and intuitive responses, while the latter was characterized by rationality and logic (Ma et al., 2014). It aligned with the earlier description of System 1 and System 2 in the risk perception process. Additionally, it was consistent with the descriptions of the 'analytic system' and 'experiential system' in cognitive psychology and neuroscience (Slovic et al., 2004).

This dichotomy has been closely associated with risk analysis, emphasizing that the seemingly contradictory dual analytical modes are essential for understanding risk, and their roles are complementary. For instance, during the risk perception process, System 1 operated automatically and continuously provided information to System 2. System 2, as a rule, embraced the information provided by System 1 and consistently adjusted to the result of risk perception. Consequently, acknowledging both systems are widely recognized as essential for a comprehensive understanding of the risk perception process (Aven, 2018). In emergency or potential risk situations, System 1 (i.e., automatic activation) may assume a pivotal role in risk perception, as risk perception represents an immediate response to potential hazards encountered (B. G. Lee et al., 2021; Loewenstein et al., 2001). Even complex studies by neuroscientists reinforced this perspective

(Damasio, 1994), indicating that logical argumentation and analytical reasoning were rendered ineffective unless guided by emotion and affect.

Based on the definition of risk perception and the two underlying processes involved in perceiving risk, it can be deduced that risk perception refers to an individual's subjective sensing and understanding of various external objective risks (Sitkin & Pablo, 1992; Van der Velde et al., 1992). Due to its perception process occurring internally within the human body, it has remained challenging to observe and detect, thus possessing a high degree of concealment. It has presented significant challenges in assessing the perceived risk.

2.3 Assessment of Perceived Risk

Risk perception significantly influences decision-making and behavioral responses in dangerous or emergencies. An accurate perception and assessment of potential risks are crucial to an ensure effective and timely emergency response to prevent accidents, minimize harm, and reduce accident losses. It is the last line of defense in safeguarding against accidents (Woodcock & Au, 2013). Researchers from various academic fields, such as sociology, safety science, and ergonomics, have been increasingly focused on this issue. In this overview, we will highlight key findings from these perspectives.

2.3.1 Risk Perception in Social Science

The research on risk perception in social science can track back to the 1960s and 1970s, and the origins were the identification of a disconnect between public perceptions of risk and technical assessments of risk (Wilson et al., 2019), then focused on its influence on decision-making and actions (Rundmo, 1995; Slovic, 1987; Slovic et al., 1982). Slovic and his research team made significant contributions to risk perception.

2.3.1.1 The existences and determinants of risk perception

Social science initially focused on the existence of risk perception. It emerged as a leading area of research in the 1980s (N. Pidgeon, 1992). Scholars predominantly directed

their attention to various types of risks perceived by individuals in their daily lives (e.g., tourism (Cui et al., 2016), smoking (Liu & Hsieh, 1995), food safety (Hansen et al., 2003), natural disasters (Gierlach et al., 2010; Ho et al., 2008; Västfjäll et al., 2008), environmental risks (Marris et al., 1996; Steg & Sievers, 2000), as well as industrial risks (B. Choi et al., 2019; Earle & Lindell, 1984; Lopez Vazquez, 2001). It is worth noting that within the industrial risks, such as explosions and toxic substance leaks, there is a particular interest when these incidents extend beyond the factory premises and significantly impact the surrounding residents, a type of risk studied in the social science field. Research findings indicated that intuitive risk perception emphasized the characteristics of potential consequences rather than the likelihood of experiencing those consequences (Wilson et al., 2019). It could be the possible reason for the bias of risk perception.

Various factors can influence the process and final assessment result of perceived risk. Perceived risk, distinct from the 'real risk', often exhibits a combination of being frightening and unknown (Slovic et al., 1985), and correlates with emotions, such as fear and anger (Sandman, 1989). These observations suggested a closer relationship between the perceived "risk" and subjective interpretative modes (Michalsen, 2003). It was precisely due to the dual primary dimensions of risk perception - cognitive and emotional dimensions- that people's perceptions of risk were susceptible to influence by various factors (Michalsen, 2003), including culture (Knuth et al., 2014), gender (Harris & Jenkins, 2006; Herrero-Fernandez et al., 2016), knowledge (Johnson, 1993), emotion (Sjöberg, 2007), experience (Ho et al., 2008), and social media (Agyeiwaah et al., 2021; Dyer & Kolic, 2020). For example, the public tended to underestimate risks in familiar, everyday situations, but there were occasions when they overestimated risks under the influence of emotions (Haluik, 2016). Factors influencing risk perception have long been a central focus of research, as both risk perception and related behaviors have the potential

to magnify the social, political, and economic impacts of risks beyond their immediate consequences (Burns & Slovic, 2012).

2.3.1.2 The assessment paradigm of risk perception

Given the significance of risk perception, achieving an accurate assessment of perceived risk becomes a critical endeavor. Currently, the most popular and widely employed approach is survey-based measurement. The psychometric paradigm, serving as a quintessential representative for quantifying perceived risks through questionnaire-based assessments, was widely applied (Marris et al., 1998; Slovic, 1990).

The psychometric paradigm played a significant role in numerous studies that quantitatively assess and compare risk perception (Alrawad et al., 2022; Bronfman et al., 2008; Cha, 2000; McCourt, 1999; Siegrist et al., 2005; Sjöberg, 2006; Slovic, 1992; Wong & Yang, 2023). It can measure human responses when confronting risk with the help of appropriate design of survey instruments. Furthermore, it employed numerical rating scales to represent their natural feelings for each question. Two essential assumptions in this approach were that risk was inherently subjective and people could provide meaningful answers to questions included in the questionnaire (Slovic, 1990). The advantage of the psychometric paradigm was that it placed a greater importance on subjective judgment (Renn, 1992). However, the simplistic attitude or questionnaire techniques to explicitly predict the dimensions of risk perception differences may be feasible at a public level but are limited to individual-level measurement and comparison (Pidgeon, 1998).

These findings in social science provide a strong foundation for effective risk communication, decision-making, policy formulation, and social-level risk management. Additionally, they have arguably prompted individuals to transition from a narrow view of risk, solely defined as the probability of an event occurring, toward a more

comprehensive framework. This framework facilitates a deeper understanding of how social dynamics depict, communicate, and influence risks (Pidgeon, 1998).

2.3.2 Risk Perception in Safety Science

In the field of safety science, human factors were considered a crucial component of systems, assessed from a system theory perspective. This assessment naturally included the consideration of risk perception. In safety science, the primary focus of risk centered around understanding human interactions with technological hazards (Marshall, 2020). However, another aspect of risk assessment is the subjective risk perceptions carried by workers directly involved in hazardous situations (Mearns & Flin, 1995).

2.3.2.1 The existence of risk perception in safety science

As early as the 1990s, scholars conducted dedicated research on risk perception and published their findings in the *Journal of Safety Science* (Flin & Mearns, 1994; Okrent, 1998; Rundmo, 1992a, 1992b, 1996). These studies initially confirmed the existence of risk perception. Risk perception was regarded as an individual's subjective assessment of objective risk sources (Committee, 1980; Rundmo, 1992b). For example, subjectivity was evident as individuals tended to weigh the consequences more heavily and be less influenced by probabilities in their risk perception (Mearns & Flin, 1995). Simultaneously, the connection between risk perception and system safety has been extensively studied in various fields, such as construction (B. Choi et al., 2019; Hasanzadeh et al., 2018; B. G. Lee et al., 2021), traffic (Cristea & Delhomme, 2016; Harbeck & Glendon, 2018; Herrero-Fernandez et al., 2016; Lu et al., 2013), and aviation (Ji et al., 2011).

Related studies have recognized the relationship between risk perception and safety-related actions. In the early 1990s, research identified that defects in risk perception were causal factors in occupational accidents. Furthermore, analyzing possible reasons for defects has revealed two main factors: the incorrect estimation of objective risk and over-estimation of one's own abilities (Rundmo, 1992b, 1996). In the subsequent studies, the

relationship between risk perception and safety behavior was evidenced (Fyhri & Phillips, 2013; Tixier et al., 2014; Xia et al., 2017). Even in questionnaire-based research within a construction context, rational assessment was found to have a limited impact on safety behavior (Xia et al., 2017). For instance, research conducted to explore the effects of not wearing helmets (Fyhri & Phillips, 2013) found that individuals who frequently wear helmets decreased in cycling speed when not wearing helmets. As for infrequent users, they showed no significant change in behavior. Within the construction context, when workers underestimate risks, they are inclined to conduct risk-taking behaviors, such as operating equipment without securing seatbelts or walking below the suspended loads (Patel & Jha, 2015; Tixier et al., 2014). In addition, drivers with a better ability to perceive risk showed a lower likelihood of crashing (Horswill et al., 2017). These studies inferred a direct relationship between risk perception and behavior. It can be said that the rationale for investigating biases in risk perception lies in the scholarly recognition of risk perception as a significant intrinsic factor influencing safety behaviors (Wang et al., 2016; Xia et al., 2017).

2.3.2.2 The biases and influence factors of risk perception in safety science

Just as perceived risk biases existed in the social sciences, similar biases were also found in the industrial domain. The divergence in perceived risk was observed among individuals at different levels within an organization and among the employees. For example, when it came to slipping hazards, drilling personnel often saw a higher level of risk during their operations, whereas management personnel tended to view this risk as negligible (Mearns & Flin, 1995). Additionally, research has demonstrated that, in general, employees were more aware of the risks they encountered in the workplace. These perceptions tended to align more accurately with quantitative risk assessment calculations and accident statistics compared to assessments made by managers. A similar pattern was observed in another study, where employees were more likely than managers

to consider their workplace as "somewhat unsafe". This difference was statistically significant (Hon et al., 2023). This discrepancy has also been observed among personnel who have experienced an accident and those who have not (Rundmo, 1995). These biases, observed in workers when perceiving risks, were recognized as contributing factors to accidents and often fell under the category of psychosocial risks.

Researchers explored factors influencing risk perception, including some unique to particular industrial fields. Alongside the well-documented factors like gender, age, education, and cultural differences, discipline-specific characteristics emerged. As early as 1992, a study employed a questionnaire survey to explore the affected factors to risk perception on offshore petroleum. Three dimensions- evaluations of disasters and major accidents, ordinary occupational accidents and post-accident measures displayed a direct connection with perceived risk. For example, more confident workers of post-accident measures often displayed lower risk for potential hazards. As further research progresses, an increasing number of influential factors have been discovered. For instance, one such factor was Inattentional Blindness (IB), where an individual's fails to notice or recognize a visual object or event due to a lack of active attention in a given situation. Studies showed that IB accounted for a significant 50% of errors in risk perception (Park et al., 2022). The level of risk perception was also closely tied to safety awareness and knowledge of the employees (Liu et al., 2021). A positive attitude enhances the effectiveness of safety knowledge (Jiang et al., 2015; Rundmo, 1992b). This positive safety culture extends to organizational culture, where a positive culture has been correlated with improved hazard recognition and risk perception among employees (Mohamed, 2002; Pandit et al., 2019). Additionally, individuals with higher scientific reasoning abilities (Siegrist & Árvai, 2020), injury experiences (Shin et al., 2014) and less distraction (Namian et al., 2018) tended to have better risk perception skills.

Regarding personal characteristics, individuals with higher emotional stability or lower anxiety levels (Bouyer et al., 2001) tend to perceive risks as lower (Chauvin et al., 2007).

2.3.2.3 The significance of risk perception in safety management

The existence of risk perception has been confirmed in safety science. The crucial significance of risk perception for system safety is the assumption that a ‘correct’ perception of risk can effectively prevent accidents (Committee, 1980). In other words, when an individual’s subjective estimation aligns with the “real” risky situation, appropriate and correct human behaviors are taken to eliminate or mitigate risk, ultimately preventing accidents. Conversely, when the subjective evaluation of potential risk sources does not correspond to the actual situation, an accident is more likely to occur (B. Choi et al., 2019; Byungjoo Choi, Gaang Lee, et al., 2019; B. G. Lee et al., 2021; Rundmo, 1992b, 1996). The conceptual framework depicting the process of risk perception in accident prevention is shown in Figure 2.1.



Figure 2.1: Conceptual safety management process (Namian et al., 2016)

Like proper hazard recognition and assessment, accurate risk perception is also regarded as fundamental to risk management. Therefore, the assessment of perceived risks should be integrated into the safety management system.

Risk perception exerts influence on the effective implementation of safety management regulations and measures (Flin & Mearns, 1994; Rundmo, 1992b). A safety regime focused on reducing risks should be sensitive to differences in risk perception (Nicholas, 2006). As workers are responsible for implementing safety management

regulations. If workers subjectively perceive that the risk associated with a particular job is low, even though it may be high, they may view management systems as an additional operational burden. Consequently, they may be unwilling to comply with these regulations. In such circumstances, the system essentially becomes a mere formality. Similarly, when workers underestimate the risk, it becomes challenging to adhere to safety practices (Zhang et al., 2015).

Risk perception plays a prominent role in risk identification, personnel selection, safety training and other aspects. The bias in risk perception can offer safety managers a more comprehensive understanding of risk information. For instance, a biased perception of risk can lead to misjudgments of potentially hazardous risk sources, resulting in inappropriate decisions and actions in everyday occupational accidents and catastrophic events (Rundmo, 1996). A study (Jing et al., 2023) revealed that risk perception can predict risky driving behaviors. Moreover, those workers who estimate risks as low are likely to have high accident rates (HSC, 1993). Therefore, identifying individuals or specific occupational settings with biased risk perception aided safety professionals in uncovering unforeseen risks and provided a foundation for revising existing safety procedures (Gürçanlı et al., 2015). Identifying individuals with accurate risk perception can facilitate the selection of more suitable candidates for high-risk job positions. Additionally, recognizing individuals with biased risk perception, and subsequent causal investigations into the causes can assist managers in tailoring more targeted safety training programs.

2.3.2.4 The assessment of perceived risk in safety science

Limited research focused on quantifying an individual's risk perception, even though it has become increasingly important in safety science. Traditionally, engineering risk assessment used the combination of probability and consequence to depict specific scenario risks (Aven & Kristensen, 2005). These methods assumed the existence of

objective risks and were used as tools to estimate these objective risks (Aven & Kristensen, 2005). These approaches formed the foundation of safety management, with countermeasures developed based on these assessments to enhance overall system safety (Aven & Kristensen, 2005).

(a) The commonly used engineering risk assessment methods

Risk assessment as a cornerstone of safety management practices can date back to the 1970s and 1980s (Aven, 2016; MacDonald, 2006). These methods were commonly regarded as quantitative methods. They involved several statistical and probabilistic tools, including Failure Mode and Effects Analysis (FMEA) (developed in the 1940s) and Hazard and Operability Analysis (HAZOP) (developed in 1963) (Swuste et al., 2014), Fault Tree Analysis (FTA) (Stamatelatos et al., 2002), Event Tree Analysis (ETA), and Bayesian networks (Curtis, 2012) to quantify risk (Waring, 2015). These techniques were commonly utilized in diverse industrial sectors. For example, FTA was conducted to analyze fall-from-height accidents in construction (Zermane et al., 2022), considering factors like an improper position for the task, improper placement and inadequate or improper Personal Protective Equipment (PPE). The proportion of unsafe behaviors in accident causation could be determined through statistical data, e.g., an 85.93% incidence rate of not wearing personal protective equipment. Results from these assessment methods were often presented numerically. Safety professionals could then compare these results with the expected or acceptable risk level and decide on appropriate measures, such as avoidance, reduction, transfer, or retention (Aven, 2016; La Fata et al., 2021).

Data used for risk assessment generally came from machine operations during the production process or historical accident records (Fung et al., 2010). However, the rarity of accidents posed challenges for data collection (Zio, 2018). Ensuring the accuracy of assessment results requires high-reliability, comprehensive, and necessary data (Li, 2014). For over 40 years, probabilistic analysis has been the basis for quantifying risk (Rechard,

1999, 2000). Another qualitative method for risk assessment involved likelihood using categories such as extreme, major, moderate, minor, and incidental, while the impact scale ranged from frequent, likely, possible, and unlikely to rare. For example, the 'extreme' impact meant significant injuries or fatalities to employees or third parties. In contrast, the 'likely' level of likelihood refers to an occurrence once every 2 to 25 years (Curtis, 2012). However, the choice of which scale to represent actual risk was subjective and situation-dependent. Experts frequently furnished assessment results grounded in their expertise and understanding of a specific area (Fung et al., 2010).

Traditional risk assessment often prioritized the dangers posed by machines over human factors (Pasman et al., 2013) despite the significant contribution of risk perception to accidents. One possible explanation is that complex and variable nature of humans made it challenging (Moura et al., 2017). Besides, many parameters of technical assessment models were typically estimated using historical data or expert opinions. Historical data about humans may suffer from inadequacies or unavailability (Gürçanlı & Müngen, 2009).

(b) The differences between engineering risk and perceived risk assessment

The assessing paradigms of engineering methods and perceived risk were fundamentally different and could not be applied to each other's methodologies. Risk perception known as System 1 thinking, which involves automatic, rapid, instinctual, and emotion-influenced risk assessment, has often been overlooked in these assessments (Aven, 2018; Okrent, 1998). The traditional engineering risk assessment methods are based on what is known as System 2 or "analytical system" thinking, characterized by a slower, more logical, and deliberative approach to evaluation (Epstein, 1994). The different characteristics of risk make it a matter of perception rather than anything equating to a measure (MacDonald, 2006). They were exact opposite ways to qualify risk. That was because the entities responsible for conducting risk assessments and the

processes involved differed significantly. Safety professionals and engineering experts typically conduct engineering risk assessments, following established paradigms and leveraging specialized knowledge. Conversely, the process of risk perception focuses on an individual level in hazardous contexts, reflecting their intuitive perceptions and comprehension of various objective risks according to their knowledge and experience (Flin & Mearns, 1994). It inherently incorporated subjectivity (Pidgeon et al., 1992) and was highly susceptible to various internal and external influences (Flin & Mearns, 1994; Inouye, 2014; Liu et al., 2023; Okrent, 1998) rendering it prone to biases.

(c) The application of survey-based methods in assessing perceived risk

Additionally, survey-based methods were utilized to measure perceived risk based on System 1 thinking (Carriço et al., 2015; Flin & Mearns, 1994; Man et al., 2020; Zsido et al., 2020). These methods involved administering questionnaires to participants after they had experienced simulated hazardous situations, aiming to gather scores that reflected their subjective feelings. For instance, a questionnaire was used to measure risk perception in risky situations, with questions such as "I feel that I would be able to stay calm and capable of acting even in the middle of a panicking crowd" and "In an emergency, I can easily exclude disturbing stimuli" (Zsido et al., 2020). However, the answers to these questions may not correspond accurately with participants' actual responses in real emergencies. Moreover, relying on post-hoc measurements can hinder participants' poor memory, leading to a degradation in objectivity and reliability (B. G. Lee et al., 2021). Therefore, it was worth noting that the assessment methods for System 1 still depended on surveys or questionnaires, a common practice in social sciences, and continued to involve retrospective data collection.

Hence, there is a critical need to move beyond the technical-centric approach of engineering risk assessment and instead redirect our focus toward creating a fresh paradigm for risk perception. This new paradigm should revolve around operational

personnel's perceived risk and form the foundational basis for developing safety technologies and management measures (Krige, 1997). Such approaches can facilitate a transformation of existing safety management models and strategies, transitioning from a reliance solely on assessments conducted by technical experts to a more integrated risk assessment that combines the insights of both technical experts and frontline workers (Aven, 2018; Renn, 1998; Skjong & Wentworth, 2001). This transformation expanded the scope of risk assessment, providing a more comprehensive perspective and abundant decision-supporting information. It seamlessly aligned with the principles of Safety II, which considered humans as essential resources in ensuring system safety and emphasized the dynamic interdependence of humans-machines-environment within the system (Aven. & Terje., 2022; Provan et al., 2020; Wahl et al., 2020). It is fundamentally different from the engineering risk assessment paradigm of 'Safety I' which minimize system variability for risk control (Aven. & Terje., 2022; Hollnagel, 2018; Provan et al., 2020).

2.3.3 The Significant Role of Risk Perception in Emergency

The primary driving force behind the growing interest in risk perception lies in its profound impact on decision-making and behavior in emergency. Figure 2.2 illustrates the general process of risk perception. It should be noted that the entire process, from the influencing factors of risk perception to the final behaviors, is limited to individuals within the production system of safety science. It does not encompass research within the field of social sciences involving the public. Therefore, some influencing factors, such as culture, nationality, or social media, are not considered.

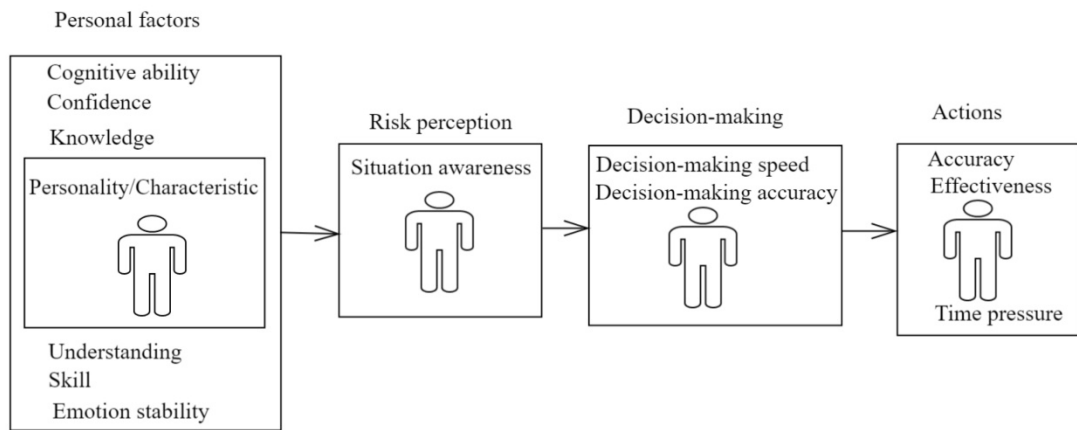


Figure 2.2: The process of how humans deal with emergency situations

Based on Figure 2.2, it was evident that the final box represented observable behaviors, and the ability to exhibit correct behavioral responses in emergencies depended on the accuracy, effectiveness, and timeliness of those behaviors (Xu et al., 2018), which were determined by the preceding step of decision-making. However, the decision-making was determined by the speed and accuracy of the decision (Flin et al., 2008), which, in turn, was influenced by perceived risk (Villa et al., 2016). If the perceived risk exceeds expectations, it may induce personal panic and hinder the ability to make appropriate decisions and take practical actions. Conversely, if the perceived risk fell within an individual's acceptable range, they would likely have more confidence in coping with the risk and making correct decisions and actions to mitigate or eliminate it. The perception of risk was influenced by situational awareness, a critical cognitive skill that significantly impacts success in emergencies at the "sharp end" (Bagley et al., 2023).

The perception process can vary among individuals, and these differences may stem from certain personal factors that can be considered as the underlying sources of performance variations. As depicted in Figure 2.2, personal factors were further categorized into two classes. The first type comprises innate individual differences, such as personality traits, which were generally difficult to change by environmental factors. For example, individuals with a disposition for fear tended to make relatively pessimistic

risk assessments (Lerner & Keltner, 2000), while highly anxious individuals tended to overestimate risks. The second type includes individual differences that can be improved or modified through training and experience, such as cognitive abilities, knowledge and understanding of risks (Jennings, 2020), and the confidence and skills required to manage risks. For instance, experienced drivers exhibited a stronger sense of control during the risk perception process (Windsor et al., 2008). These factors interacted with each other, collectively influencing risk perception and ultimately impacting an individual's performance in abnormal situations (Byungjoo Choi, Gaang Lee, et al., 2019).

The significant role of risk perception set its importance as a precursor to decision-making and subsequent behavioral responses. It can be regarded as the final psychological trigger behind the decision (Manar et al., 2019). Therefore, conducting comprehensive research on risk perception is essential to enhance suitable outcomes, especially in risk and emergency situations. However, it can also be inferred that various factors affecting perceived risk exist, thus presenting significant challenges in researching risk perception.

2.3.4 Risk Perception in Ergonomics

In the field of ergonomics, the primary goal is to enhance the interaction between systems and humans (Kazemi & Lee, 2023). Human factors, a central component, were usually assessed through cognition and reliability evaluations rather than focusing on describing, quantifying, and mitigating their impact on system safety, as was common in safety science. The former focused on intrinsic aspects of human cognition, aiming to improve human performance by elevating cognitive abilities (Fista et al., 2019). Conversely, the latter emphasized observing external behaviors and manifestations of individuals, seeking to enhance workforce reliability (Kirwan, 1992). Both disciplines shared the goal of enabling individuals to perform their tasks effectively while ensuring safety and well-being.

2.3.4.1 Cognitive assessment and its comparison with risk perception

Self-reported questionnaires and objective physiological evaluations were usually used to assess cognitive ability. Cognitive processes served as essential determinants of subsequent behavioral performance (Choi & Seong, 2020), a correlation consistently validated by several related research (Bagley et al., 2023; Gonzalez et al., 2005; Irwin et al., 2023). The practice of assessing cognitive ability dates back to 1987, as discussed in the book "Risk and Decisions" on pages 87-107 (Singleton & Hovden, 1987). However, its execution has always been challenging due to the complicated cognitive processes and influenced factors like moods, emotions, and motivations.

Currently, two primary methodologies are used to assess cognitive ability: subjective self-reported questionnaire (Fista et al., 2019) and objective physiological evaluation (Bačić & Henry, 2022; Digiesi et al., 2020; Minkley et al., 2021). Subjective methods commonly used questionnaires, such as NASA-TLX and subjective workload assessment (Digiesi et al., 2020; Widianti et al., 2017; Zimmer et al., 2019). For physiologic measurements, indicators such as Electrodermal Activity (EDA) (Armougum et al., 2020; Baldauf et al., 2009; Golmohammadi et al., 2022), Heart Rate Variability (HRV) (Digiesi et al., 2020; Grassmann et al., 2017; Tjolleng et al., 2017; Zhu et al., 2022), eye-tracking (Biondi et al., 2023; Čegovnik et al., 2018; Dehais et al., 2012; Devlin et al., 2022) and EEG (Bernhardt et al., 2019; Morton et al., 2022; Xu et al., 2023) were often employed to assess cognitive process or workload. This approach involved collecting and tracking individuals' neurophysiological signals to reflect underlying cognitive processes (Bačić & Henry, 2022).

While cognitive assessment in safety science has traditionally focused on human factors rather than risk perception, these two shared connections and distinctions. Similarities existed in that cognitive assessment and risk perception centered around individuals, relying on cognitive processes and closely intertwining with decision-making.

However, their primary focuses diverged. Cognitive assessment primarily delved into individuals' cognitive processes and workload across various tasks, aiming to align human-machine interface designs more effectively with human cognitive capabilities. On the other hand, risk perception is concerned with how individuals perceive and evaluate potential risks in urgent or risky situations, providing a foundation for subsequent decision-making and actions. The physiological measurement of cognitive assessment provided a new perspective for evaluating perceived risk.

2.3.4.2 Human reliability assessment and its comparison with risk perception

Human Reliability Assessment (HRA) strengthened the identification of specific instances of human errors and their contribution to accidents. Traditional approaches for dealing with human errors involved HRA techniques (Iqbal et al., 2020). HRA aimed to improve system performance by identifying, predicting, and reducing human errors (Kirwan, 1994; Kirwan & James, 1989). The earliest studies on HRA were conducted in the 1970s in Nuclear Power Plants (NPPs) (La Fata et al., 2023). The first classification and description of HRA methods was provided by Bell and Holroyd (Bell & Holroyd, 2009), who summarized 72 potential tools related to human reliability. Current studies have developed three generations of human error calculation models.

(a) Three generation models of HRA

In the first-generation models (e.g., Technique for Human Error Rate Prediction - THERP and Success Likelihood Index Method - SLIM) (Embrey et al., 1984; Swain & Guttman, 1983), humans were treated as elements similar to machines, and human error probabilities and operational human errors calculated based on their actual work or simulated data (Munger et al., 1962).

In the second-generation models (e.g., Cognitive Reliability and Error Analysis Method-CREAM and the Standardized Plant Analysis Risk Human Reliability Analysis - SPAR-H) (Gertman et al., 2005; Hollnagel, 1998), humans were no longer viewed as

simple elements similar to a machine. More attention was given to their subjectivity, organizational influences, and cognitive processes, focusing on the root causes of human error, such as perception, thinking, memory, decision-making, and action strategy before human errors occur (Hollnagel, 1998). Cognitive skills and technical knowledge were considered vital for safe and effective work performance (Irwin et al., 2023). In the absence of objective failure data, expert judgment was often used to predict human error probabilities (Deacon et al., 2010; DiMattia et al., 2005), but these approaches relied on subjective estimates rather than objective measurement of cognitive behaviors (Kodappully et al., 2016).

The third-generation models, known as Dynamic HRA methods, incorporated the variability of human behavior across different periods, task stages, and work environments. These models treated individuals' states and behaviors as dynamic processes, enabling the modelling and prediction of human reliability in diverse contexts. However, biases that arise during risk assessment workshops can lead to underestimating the level of risk, which can be hazardous and challenging to prevent or mitigate (Bagley et al., 2023). Currently, or more precisely, since 2005, the interrelationships and dependencies of human performance factors (Bevilacqua & Ciarapica, 2018), the specific contexts in which human errors occur, and the demands on human factors under dynamic or abnormal conditions have become research topics (Woodcock & Au, 2013). Regardless of the specific model employed, the data used for calculating human errors depends predominantly on historical data collection, expert assessments, or simulations (Pate-Cornell & Murphy, 1996). Moreover, the dynamic characteristic of humans has paid more focus.

(b) The differences between HRA and risk perception

The assessment of human reliability and risk perception exhibits distinctions and interrelations. Firstly, a primary disparity lies in their respective focal points. Human

reliability assessment primarily centers on the reliability and consistency of individuals in their work or tasks. In contrast, risk perception focuses on how individuals perceive and evaluate potential risks or hazardous situations. Secondly, their objectives diverge as well. Human reliability assessment aims to ensure that individuals can perform tasks as expected to reduce errors and accidents. Conversely, risk perception aims to ensure individuals correctly identify potential risks, preparing them for decision-making and action. However, certain connections exist between the two, such as the fact that an individual's reliability can impact their perception of potential risks and ability to respond effectively. In simpler terms, both shared the same aim to realize system safety in different realization ways.

2.3.5 Comparison of Different Risk Perception Assessment Paradigms

Although the psychometric paradigm and survey-based methods have been widely employed to assess subjective risk, it is essential to recognize their potential limitations, including recall bias, subjectivity and predictive capabilities. Furthermore, employing engineering risk assessment to represent perceived risk is constrained by the behaviors of machines and external factors. As a result, measuring risk perception, with its subjective and opaque nature, presents significant challenges in developing an objective and non-intrusive measurement to assess them. However, the physiological measurements of the cognitive process provide a new perspective to measure perceived risk.

Bio-signals offered a better way to measure the risk perception process, specifically System 1 responses. Utilizing physiological measurements has advantages in evaluating subjective risk assessment. First, physiological measurements are non-invasive and do not cause significant psychological discomfort or resistance in subjects (Zhai & Barreto, 2006). Second, the currently prevalent physiological measurement methods, such as wristbands or wireless biosensors, do not interfere with the subjects' perception of risk (B. G. Lee et al., 2021). Third, physiological signals are generated concurrently with the

subject's risk perception process. It means that the risk subjects perceive triggers physiological fluctuations, which are synchronously transmitted and recorded. Therefore, the natural physiological responses in the human body during perceiving risk bring excellent facilitation and feasibility for assessing perceived risk (Byungjoo Choi, Gaang Lee, et al., 2019; B. G. Lee et al., 2021).

There has been limited research on detecting variations among risk situations with slight divergences in their degree of riskiness. However, significant changes in physiological signals during the perception of different risk situations hold significant potential. These changes include increased cardiovascular system activity, alterations in sweat gland nerve control, and temperature regulation (Budidha & Kyriacou, 2019; Dettmers et al., 1993; Nitzan et al., 1998). Multiple studies have investigated the fluctuations in physiological signals across various risk scenarios, uncovering notable disparities. For instance, the strength of P200 and LPP components in EEG recordings have been found to increase in higher-risk situations. Other physiological indicators, such as Heart Rate (HR) (Rendon-Velez et al., 2016), EDA (Barnard & Chapman, 2016; Tagliabue & Sarlo, 2015), pupil size, and the number of saccades (Charlton et al., 2014) have also shown increases with higher levels of perceived risk. Indeed, most of these significant disparities in physiological signals have been observed between high and low situations. These findings suggested the potential utilization of individuals' physiological responses to evaluate risk at the personal level (B. Choi et al., 2019). However, further research is necessary to determine whether significant differences exist among risk levels that are more closely related, because these differences are essential for the subsequent and successful application of physiological responses.

Further research was needed to explore the hidden neuroregulatory mechanisms within physiological changes during the risk perception process. This effort can comprehensively understand the physiological responses when individuals perceive risk.

As mentioned earlier, the process of risk perception arises from the interaction between an individual's cognition, the external environment, and task operations. An individual's automatic and unconscious response to danger (referred to as the System 1 response) triggers the Sympathetic Nervous System (SNS) of the Autonomic Nervous System (ANS). SNS usually showed an increased change in risk situations. However, there is also the possibility that individuals perceive a high level of risk that exceeds their coping capacity. It may activate the Parasympathetic Nervous System (PNS) of ANS and decrease it, resulting in a psychological need for self-protection and avoidance of harm. These represented the current consensus on neuro-regulation in human risk perception. However, it remained unclear whether the SNS and PNS continued to exhibit enhancement and attenuation in situations where risk magnitude closely matched.

Research aimed at establishing models for identifying perceived risk based on physiological responses still needs to be completed. However, there are ongoing efforts in this area. One study assessed drivers' perceived risk (Ping et al., 2018), while another two focused on construction workers (Byungjoo Choi, Gaang Lee, et al., 2019; B. G. Lee et al., 2021). They were all in the context of technological risk.

2.4 Physiological and Emotional Responses in Risk Perception

By capturing an individual's real-time physiological responses to hazards, wearable devices offer a promising means to gain more accurate insights into personal risk perception (B. Choi et al., 2019). Comprehending the synchronous accompanying physiological and psychological changes and the mechanisms underlying physiological activation is essential. Therefore, the following section provides a detailed introduction to several commonly utilized physiological signals and their changes in the process of perceived risk.

2.4.1 Electrodermal Responses

2.4.1.1 The commonly used indicators of EDA

EDA was defined as the fluctuations in the electrical properties of the skin that result from sweat secretion (Benedek & Kaernbach, 2010). It was measured non-invasively by continuously monitoring changes in skin conductance or potential via two electrodes that applied a low, constant voltage to the skin (Lee et al., 2020). Electrodermal sensors are typically positioned at the areas of the highest sweat gland density, such as the hands and feet (Yang & Liu, 2014). The signal can be decomposed into a tonic, slow-changing component (i.e., electrodermal level-EDL) and a phasic, rapid changing component (i.e., electrodermal response-EDR) (Braithwaite et al., 2013; B. Choi et al., 2019; Greco et al., 2016; Irwin et al., 2023). The tonic component includes slow drift of the baseline skin conductance, and the phasic component reflects short-time and immediate responses to external stimuli (Greco et al., 2016; Lee et al., 2020).

2.4.1.2 The changes of EDA during risk perception

EDA has been widely recognized as an important biomarker for assessing risk perception (B. Choi et al., 2019). Two key factors bolster this assertion. Firstly, prior research has demonstrated the equivalence of skin conductance changes in simulated and real-world settings, possibly attributable to the similar emotional responses elicited by risk factors in both conditions (Watts & Quimby, 1979). Secondly, EDA distinguishes itself by exclusively reflecting the sympathetic branch of the autonomic nervous system (Boucsein et al., 2012; Braithwaite et al., 2013; Picard et al., 2016), remaining unaffected by PNS. When individuals perceive risk, external stimuli activate SNS, subsequently triggering postganglionic sudomotor fibers that innervate the sweat glands (Nishiyama et al., 2001). The sweat secreted by these sweat glands can lead to changes in the skin surface resistance, which can be measured using skin conductance sensors (Boucsein et al., 2012).

EDR, compared with EDL, is more sensitive to representing perceived risk. EDR consistently exhibits statistically significant increases in high-risk contexts, possibly because its amplitude reflects arousal levels (Dawson, 2007). This phenomenon has been corroborated across various risk scenarios, including construction, traffic driving (Liang & Lin, 2018), and aviation safety (Gao & Wang, 2020). Furthermore, irrespective of the specific nature of the high-risk situation, whether it involves hazardous working conditions (Byungjoo Choi, Gaang Lee, et al., 2019), varying levels of accident risk (Distefano et al., 2022), or stages of heightened danger presentation (Barnard & Chapman, 2016), EDR consistently showed an increasing trend (Barnard & Chapman, 2016; B. Choi et al., 2019; Gao & Wang, 2020; Liang & Lin, 2018). However, concerning EDL, one study indicated that it exhibited a growth trend but no significant variation in both high and low-risk contexts (B. Choi et al., 2019), while much of the research exclusively analyzed EDR and neglected EDL (Giagloglou et al., 2019; Liang & Lin, 2018).

2.4.2 Cardiac Responses

The electrocardiogram (ECG) is a method used to monitor the activity of the heart (Ahmed et al., 2022). It is also non-invasive and involves placing small electrodes on the skin of the chest, arms, and legs to detect the electrical impulses generated by the heart during each heartbeat (Magnon et al., 2022; Shaffer et al., 2014; Sztajzel, 2004). The neural regulation of the heart is more complex than that of EDA because it is simultaneously regulated by both SNS and PNS, which have antagonistic effects within ANS (Berntson & Cacioppo, 1999; Berntson et al., 1997; Malik, 1996; Sztajzel, 2004). Importantly, the PNS takes less time (<1s) to exert its effect compared to the SNS, which takes more time (>5s) (Nunan et al., 2010). Furthermore, PNS stimulation only affects one or two heartbeats after its onset, while the effects of SNS stimulation can last for 5-10 seconds and continue to affect heart activities (Shaffer et al., 2014).

2.4.2.1 The commonly used indicators of HRV

Heart rate variability (HRV) has been widely used for quantifying the activity of the autonomic nervous system. HRV can be typically decomposed into three categories: time-domain, frequency-domain and nonlinear indicators. Time-domain indicators include Heart Rate (HR), Inter-Beat Interval (IBI), and others. Frequency-domain indicators include ultra-low-frequency (ULF, below 0.0033Hz), very-low-frequency (VLF, 0.0033 and 0.04Hz), low frequency (LF, 0.04 and 0.15Hz), high frequency (HF, 0.15 to 0.4Hz) bands and LF/HF ratio. The nonlinear indicators, such as SD1, SD2, A++ and B--, can be obtained by The Poincare plot. This plot is a visual tool to represent changes between RR intervals to assess the state of the heart (Kanjo et al., 2018), and the shape of the RR interval distribution typically shows an elliptical pattern (Makivic & Bauer, 2017). Different physiological indices usually represent various information about SNS and PNS. The definition and physiological significance of HRV indicators are detailed in Table 2.1.

2.4.2.2 The changes of HRV during risk perception

HRV is also commonly used to represent risk perception, since it is frequently considered a measure of neurocardiac function that reflects heart-brain interactions and ANS dynamics (Kikuta et al., 2023; Perello-March et al., 2022; Shaffer et al., 2014). When an individual perceives a potential risk, the sympathetic division of their ANS is activated, while the parasympathetic division diminishes, thereby causing changes in heart-related activity. Previous research has also demonstrated a significant correlation between these metrics and individuals' risk perception (as evidenced by studies on HR, HRV, etc.) (Herrero-Fernandez et al., 2016; Mesken et al., 2007; Powers et al., 2008; Rubaltelli et al., 2018). Additionally, human cardiac and electrodermal measures were highly correlated in risk situations (Wilson, 2002). However, the heart rate has lower temporal sensitivity and is less suitable for assessing the effect of scenarios. Specifically,

the mean inter-beat interval is 0.75 s, and it takes at least several beats to detect a change in heart rate (Jorna, 1992; Rowe et al., 1998).

Heart activity is simultaneously regulated by both SNS and PNS, leading to some inconsistent findings when exploring the relationship between HRV and risk perception. For example, HR was found to be higher in more hazardous conditions (Doorley et al., 2015; Kikuta et al., 2023; Park, 2009; Prell et al., 2020; Tagliabue & Sarlo, 2015), but it showed a decline in driving crashes (Barnard & Chapman, 2016). Similar decreased results of the time domain index SDNN of HRV in elderly drivers were found with the rise of risk level (Xiong & Guo, 2021). Further studies should be conducted to validate the changes in HRV and their correlation with perceived risk.

2.4.3 Eye Tracking Behaviors

The non-invasive and relatively simple installation and use of eye trackers make them appropriate for monitoring cognitive state (Matos, 2010). Eye movements have been recognized as a valuable measurement of cognitive processes since Javal's discovery in 1878 that they could reflect internal cognitive processes (Wade & Tatler, 2009). Specifically, Javal found that during reading, eye movements occur in a sequence of discrete pauses called fixations, separated by jumps called saccades. Fixations are believed to be associated with visual information processing, while vision is essentially suppressed during saccades (Wade & Tatler, 2009). Eye movements are widely regarded as a real-time window into cognition, based on the "eye-mind" hypothesis, and are therefore capable of providing dynamic information about human cognitive behavior (Cooke, 2005).

2.4.3.1 The basic eye-tracking indicators

Eye tracking technology has advanced significantly in recent years, with the development of non-intrusive, accurate, and readily available eye trackers such as Tobii, SR Research, and SMI (Kodappully et al., 2016). These devices typically use an infrared

light source directed towards the eye, with a camera recording reflection from the eyes. The resulting data is processed to extract information about pupil size, fixation, and saccades, among other parameters (Kodappully et al., 2016). To ensure accuracy, many studies using non-invasive eye-tracking methods follow a nine-point calibration process, where participants are asked to focus on a red dot appearing at predetermined screen points (Kodappully et al., 2016). Following calibration and measurements, a series of eye movement indices can be extracted through data preprocessing and algorithms (Kodappully et al., 2016). These indices can provide insights into various cognitive and perceptual processes, including attention, visual processing, decision-making and memory. For instance, pupil size is the changes in pupil size that can be indicative of changes in cognitive workload or arousal levels. Related indices commonly used are listed in Table 2.1.

2.4.3.2 The application of eye-tracking during risk perception

The application of eye-tracking behavior on risk perception mainly focused on two aspects. One is to research the relationship between risk perception and gaze behavior. For example, subjects would emphasize areas perceived as more dangerous during urban cycling (Schmidt & von Stülpnagel, 2018). Similarly, experienced drivers fixated on predefined regions with concealed risks (Pradhan et al., 2006). The other is utilizing eye-tracking technology to assist in uncovering issues in workers' risk perception processes. In the construction site context, scan paths and attention maps generated using eye-tracking technology can effectively demonstrate the risk recognition process to construction workers, aiding them in identifying previously unnoticed risks through subsequent analysis (Jeelani et al., 2018). In another study, wearable eye-tracking technology was used to analyze and compare the eye-tracking data of experts and novices in the electrical field, including metrics such as gaze fixation, count, and average fixation

duration. The analysis revealed that experts exhibited more concise risk perception trajectories and shorter total attention time (S. Li et al., 2022).

2.4.4 Skin Temperature Changes

Changes in microvascular blood volumes have been shown to affect peripheral skin temperature (ST) (Aryal et al., 2017; Garbey et al., 2007). Therefore, ST can indicate cardiac activities triggered by the stimulated sympathetic nervous system (Dias & Cunha, 2018; Kim et al., 2004). It was utilized as an input variable in the assessment of perceived risk in construction areas (B. G. Lee et al., 2021). For temperature measurements, its mean value is sufficient.

Varied changes in physiological indicators stemming from bio-signals convey different physiological significance and reflect internal human processes. Consequently, Table 2.1 provides a list of commonly used indicators for physiological measurement signals, along with their respective meanings.

Table 2.1: The common physiological indicators and their meanings

No.	Measurement	Meaning	Interpretation
EDA			
1	EDL/ μs	A slow changing component and slow drift of the baseline skin conductance.	It is considered to show the long-term stress trend in sympathetic activity (Braithwaite et al., 2013; B. Choi et al., 2019; Poh et al., 2010).
2	EDR/ μs	A short-time and immediate responses to external stimuli.	It can well reflect the short-term sympathetic response to an external stimulus (B. Choi et al., 2019; Byungjoo Choi, Gaang Lee, et al., 2019).
ECG			
3	Heart Rate (HR) /bpm	The number of times a normal person's heart beats per minute at rest (beats per minute).	It can reflect the relative activity of the sympathetic and parasympathetic systems (Sztajzel, 2004).

Table 2.1, continued

No.	Measurement	Meaning	Interpretation
4	MeanRR/ms	The time interval between two consecutive heartbeats, also known as the R-R interval.	The longer the NN interval, the slower the heart rate (more HRV). Conversely, the shorter the R-R interval, the faster the heart rate (less HRV). Higher resting heart rate and lower resting HRV usually reflect a compromised physiological state (Kemp et al., 2017).
5	SDNN/ms	The standard deviation of the heart rate-corrected N-N intervals for all beats in 24-h recordings (Shaffer et al., 2014).	It can be used to assess the overall size of HRV, reflecting the slow changes in heart rate and it is a sensitive indicator for evaluating sympathetic nervous system function (Sztajzel, 2004).
6	SDANN/ms	The standard deviation of the average NN intervals (mean heart rate) for each of the 5-min segments during a 24-h recording (Shaffer et al., 2014).	It reflects the balance of short-term regulation between the sympathetic and parasympathetic nervous systems. It is correlated with the SDNN (Shaffer et al., 2014).
7	SDSD/ms	Standard deviation of differences between successive NN intervals (Ahmed et al., 2022).	Higher SDSD value indicates greater variability between adjacent heartbeats, which is generally considered as a reflection of the balance between sympathetic and parasympathetic nervous system control (Sztajzel, 2004).
8	RMSSD/ms	The root mean square of the differences between consecutive IBI intervals for all beats (Shaffer et al., 2014).	It is the short-term sensitive indicator to assess PNS function and correlated with HF power (Kleiger et al., 2005).

Table 2.1, continued

No.	Measurement	Meaning	Interpretation
9	ULF	The percentage of ultra-low-frequency energy in the total energy (Ahmed et al., 2022).	The increase in ULF component may be related to long-term regulation of the sympathetic and parasympathetic nervous systems (Kleiger et al., 2005).
10	VLF	The percentage of very-low-frequency energy in the total energy. Power in the very low frequency range(Ahmed et al., 2022)	It reflects the long-term regulatory effects of the sympathetic and parasympathetic nervous systems (Armour, 2003).
11	LF	The percentage of low-frequency energy in the total energy (Ahmed et al., 2022).	It usually reflects the activation information of the sympathetic nervous system, but it seems to be also influenced by parasympathetic activity (Hayano & Yuda, 2019).
12	HF	The percentage of high-frequency energy in the total energy (Ahmed et al., 2022).	It usually reflects the activation information of the parasympathetic nervous systems (Kemp et al., 2017).
13	LF/HF	The ratio of low-frequency to high-frequency energy (Ahmed et al., 2022).	It usually reflects the balanced control of the autonomic nervous system, depends on the recording condition in which data is collected; it can reflect vagal, sympathetic and baroreflex mechanisms (Kemp et al., 2017).
14	SD1/ ms^2	The standard deviation of the shaort axis of the ellipse in a Poincaré plot.	It is a measure of rapid changes in the N-N interval and is therefore considered a parasympathetic index of sinus node control. Correlates highly with RMSSD and HF measures (Kemp et al., 2017; Rajendra Acharya et al., 2006).

Table 2.1, continued

No.	Measurement	Meaning	Interpretation
15	SD2/ ms^2	The standard deviation of the long axis of the ellipse in a Poincaré plot.	It represents long-term variability and is more strongly associated with sympathetic nervous system activity than with parasympathetic nervous system activity. Specifically, when SD2 decreases, sympathetic nervous system activity decrease (Kemp et al., 2017; Kemp & Quintana, 2013).
16	A++	The points in the first quadrant of a scatter plot.	Indicate an increase in the interval between two consecutive heartbeats and a decrease in heart rate, representing parasympathetic nervous system activity (Moharreri et al., 2018; Network, 2023).
17	B--	The points in the third quadrant of a scatter plot.	Indicate a decrease in the interval between two consecutive heartbeats and an increase in heart rate, representing sympathetic nervous system activity (Moharreri et al., 2018; Network, 2023).
Eye tracking			
18	Fixation duration	The length of time the eyes remain fixed on a particular area of interest.	Longer fixation durations may indicate increased attention or cognitive effort (Bjørneseth et al., 2014).
19	Saccade length	Rapid eye movements that occur between fixations.	The length of these movements can provide information about the efficiency of visual processing (Yan et al., 2019).

Table 2.1, continued

No.	Measurement	Meaning	Interpretation
20	Pupil size	The changes in pupil size	It can be indicative of changes in cognitive workload or arousal levels (Schmidt et al., 2017; Srinivasan et al., 2019).
21	Gaze duration	The total time spent fixating on a particular area of interest.	It can provide insight into the amount of attention paid to a particular stimulus (Martinez-Marquez et al., 2021).
22	Number of fixations	The number of times the eyes fixate on a particular area of interest.	It can provide information about the salience of a particular stimulus (Atik & Arslan, 2019).
23	Blink rate	The average number of blinks per unit of time.	Changes in blink rate can be indicative of changes in cognitive workload, fatigue or arousal levels (Atik & Arslan, 2019; Wang et al., 2013).

2.4.5 Emotional Responses

The study of emotional influence on risk perception originated in 1978 (Fischhoff et al., 1978). Subsequently, other scholars advanced the notions of "risk as feels" (Loewenstein et al., 2001) and the "affect heuristic" (Finucane et al., 2000). As for emotional reactions, discrete types of emotions, such as fear (Lerner & Keltner, 2000; Longin et al., 2013), anxiety (Janelle et al., 1992), and anger (Deffenbacher et al., 2003; Mesken et al., 2007) were often studied. Related research mainly focused on the influences of different emotions on risk perception (Drače & Ric, 2012; Herrero-Fernández et al., 2020; Jeon & Zhang, 2013; Lu et al., 2013; Parrott, 2017; Rundmo, 1996; Slovic & Peters, 2006; Yang & Chu, 2018). For example, anger reduces risk perception, whereas fear increases it (Lu et al., 2013). In these studies, emotions were first induced, followed by an examination of their impact on risk perception. However, the present study

focused on the naturally induced emotions during the risk perception process (Agyeiwaah et al., 2021; Dyer & Kolic, 2020; Klemm et al., 2019; Tixier et al., 2014).

2.4.5.1 The induced complicated emotions during risk perception

The presence of induced affects by risky situations has been researched mostly, but they were highly complex. When individuals were involved in urgent or dangerous situations, the evoked emotions were often not singular but rather complex, potentially encompassing fear, worry, anxiety, despair, and excitement (Hasanzadeh & De La Garza, 2020; Loewenstein et al., 2001; Slovic et al., 2004). Moreover, these emotions were dynamic and fluctuated depending on the situation's progression, such as the effectiveness of actions taken and the severity of the risk. The complexity and dynamics of these emotions pose significant challenges in their representation.

The induced complicated emotions played a pivotal role in risk perception. The initial perception of risk triggered emotions, subsequently guiding the development of more comprehensive and refined appraisals (Sjöberg, 2007). According to the risk-as-feelings hypothesis, feelings, particularly autonomic and immediate responses, played a significant role in risk perception (B. Choi et al., 2019; B. G. Lee et al., 2021; Slovic et al., 2004). Moreover, behavioral evidence (Loewenstein et al., 2001) suggested that emotional responses and cognitive evaluations were somewhat separate in the risk perception process, with emotions often influencing the outcome of risk assessments. Emotional responses guided reactions during the system 1 stage and continued to impact assessment outcomes in subsequent cognitive evaluation stages through conditioned reflexes and memory as bodily markers. Even in seemingly objective risk situations, affective reactions still hold considerable influence (Slovic & Weber, 2013).

2.4.5.2 The expression of induced emotions

Given that emotions naturally arise with risk perception, they could be as informative for risk perception as physiological responses. Unfortunately, it is regrettable that there

is currently limited research incorporating risk-induced emotions. The issue may lie in the expression of complex emotions. Fortunately, the Pleasure-Arousal-Dominance (PAD) emotional model may be introduced to express emotions. Specifically, emotions can be represented as points in a three-dimensional space defined by valence (the degree of positivity or negativity of an emotion), arousal (the intensity or activity level of an emotion), and dominance (the perceived level of control or power a person feels in the situation), as proposed by Wundt in 1922 (Wundt, 1922). The successful application of PAD (Ding, Ghazilla, et al., 2022) provides great confidence for this research. Importantly, the perception of risk is influenced by controllability, where individuals tend to perceive less risk in situations they control (Brun, 1994). Considering the meaning of the dominance dimension, it may be more appropriate for describing the feeling of whether a person can control the situation or is being controlled by emergency or risky circumstances.

2.4.6 Discussion of Physiological and Emotional Responses

Existing research has investigated the relationship between risk perception and accompanying physiological responses. Variations in risk situations lead to distinct physiological changes, often exhibiting statistically significant differences. It provides a robust foundation for further developing models to classify risk levels.

Furthermore, the utilization of multiple physiological signals to predict perceived risk may prove to be more effective. When assessing cognitive load, researchers have employed a combination of EDA and ECG (Shayesteh et al., 2023), integrated EEG, PPG, and EDA (Alrefaie et al., 2019), and combined ECG, EEG and pupil diameter (Orlandi & Brooks, 2018), all of which have yielded promising results. Likewise, in assessing construction workers' perceived risk, a combination of EDA, ECG and skin temperature (ST) was employed (B. G. Lee et al., 2021). For drivers' risk identification, a combined bio-signals of EDA, ECG, Respiration (RESP) and pupil diameter (Ding, Ghazilla, et al.,

2022) or driving actions (Ding, Raja Ghazilla, et al., 2022) were used as input data to train ML models. Perhaps the integration of multiple physiological signals can provide more comprehensive insights into the body's response in specific contexts.

2.5 Application of Machine Learning

With the emergence of Industry 4.0 in 2011 (Adem et al., 2020; Rosi et al., 2018), a series of new technologies have a chance to develop. Artificial intelligence (AI) is one of the central methods, considered a system consisting of software and hardware to imitate intelligent human behavior for decision-making (Erol, 2019). AI, or more precisely, machine learning, is gradually permeating the research field of risk assessment.

2.5.1 Application of Machine Learning on Risk Assessment

Despite approximately thirty different risk assessment technologies, the performance of traditional methods is limited due to a lack of dynamic and real-time analysis (Hegde & Rokseth, 2020). In recent years, the methods of risk assessment have transformed. Emerging machine learning (ML) methods, driven by data entirely, are rapidly developing in this field. There are dedicated review papers discussing the application of machine learning on risk assessment. Next, a brief introduction will be provided on the application of machine learning in risk assessment.

2.5.1.1 Application phase for ML

According to the International Organization for Standardization, risk assessment is separated into three phases: risk identification, risk analysis and risk evaluation (Hutchins, 2018). From the meanings of these three stages (detailed in Table 2.2), it can be observed that each subsequent stage builds upon the previous stage. Risk identification involves the identification or recognition of existing hazards. Risk analysis entails a qualitative or quantitative assessment of the magnitude of each risk based on the identified hazards. This assessment can be expressed through qualitative ratings or a combination of quantitative probabilities and consequences. Risk evaluation involves comparing the

evaluation results with the acceptable risk level established during the risk analysis stage. If the evaluation results exceed the acceptable level, the risk is considered unacceptable, and measures must be implemented to mitigate the risk.

The goals differ across three assessing stages, and the machine learning methods vary slightly. For instance, in the context of risk analysis and evaluation, it is imperative to quantify the probability of potential accidents where machine learning algorithms, such as Bayesian networks and logistic regression, can be deliberately selected. The typical examples for different phases are also listed in Table 2.2.

Table 2.2: ML application for different phases of risk assessment

No.	Risk assessment phase	Definition	Typical application
1	Risk identification	The process of identifying potential risk factors.	Hazard prediction for enterprise workplaces(Wang et al., 2018)
2	Risk analysis	The process of comprehending the nature and determine the level of risk.	Maritime Traffic Probabilistic prediction (Xiao et al., 2017)
3	Risk evaluation	The process of comparing the results of risk analysis with risk criteria to determine whether the risk and/or its magnitude is acceptable	Evaluation the risk of crashing a car into the vehicle in front(Curiel-Ramirez et al., 2019)

In the review paper (Hegde & Rokseth, 2020), the number of published papers for the three risk assessment phases are summarized in Figure 2.3.

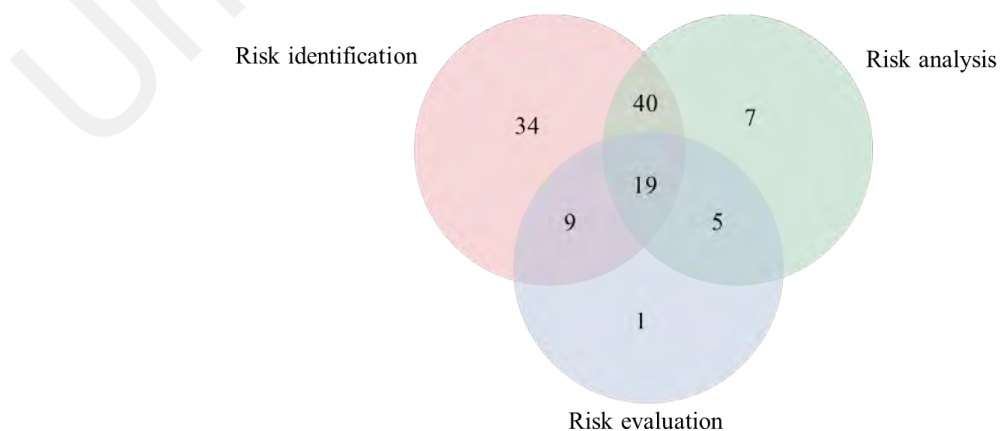


Figure 2.3: The number of papers concerning the three risk assessment phases (Hegde & Rokseth, 2020)

2.5.1.2 Data sources for ML

Three primary sources of data are used to train ML models for risk assessment: historical data, real-time data, and a combination of both. The proportions of these data in risk assessment can be seen in Figure 2.4. Historical data primarily consists of data collected from system workers, machinery, and environmental aspects. Numerical data is more common in this category, but there are also textual data sources, such as accident investigation reports. For example, 5298 raw accident reports were used as input indicators to identify the attribute combinations contributing to construction industry injuries by employing natural language processing (NLP) (Tixier et al., 2017). Real-time data, on the other hand, is collected synchronously during work and operational processes. Common examples include real-time data on human physiological and psychological states, including output data from mechanical equipment. Such as vehicle information was used as input data to conduct a real-time highway traffic condition assessment using SVM (Support Vector Machine) (Ma et al., 2009).

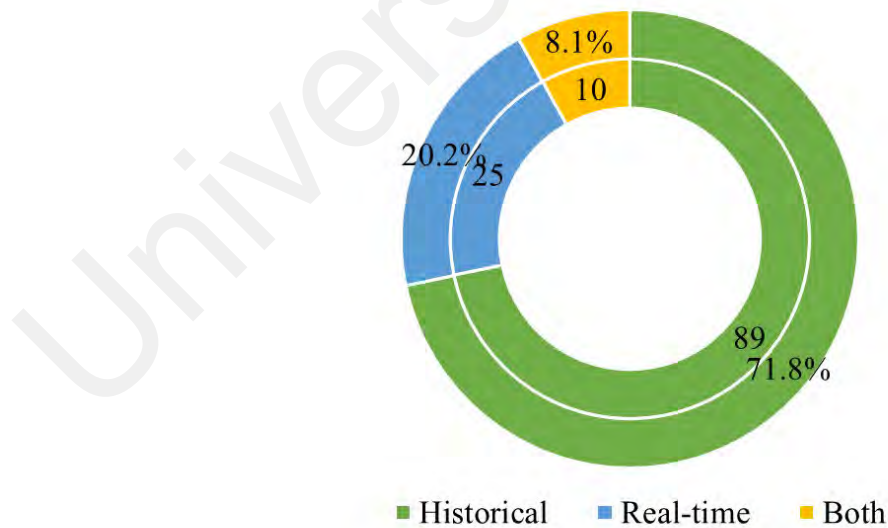


Figure 2.4: The three types of data sources to build the model (Hegde & Rokseth, 2020)

2.5.1.3 The frequency of method usage

In applying machine learning in risk assessment, the most intriguing questions would be which algorithm is the most used and which algorithm yields the best results. The answers to these two questions have also been provided. Examining the frequency of algorithm usage, as illustrated in Figure 2.5, reveals that Artificial Neural Network (ANN), Support Vector Machine (SVM), and Decision Tree (DT) are the three most frequently used algorithms. Among these, ANN, stands out due to its strong capability to capture nonlinear relationships (Bevilacqua et al., 2010) and the advantage of requiring less formal statistical training (Tu, 1996), making it the most widely used algorithm.

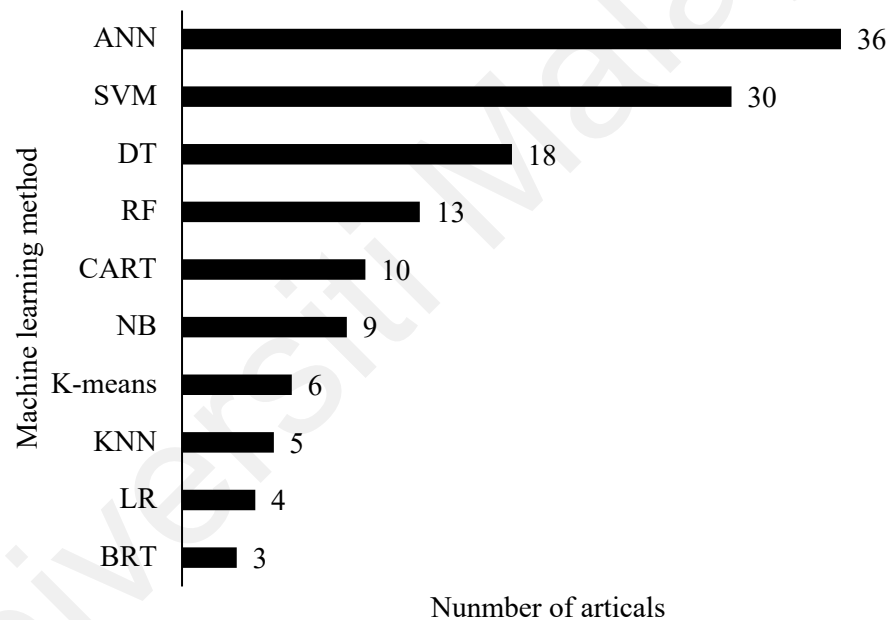


Figure 2.5: Ten frequently used machine learning methods in risk assessment
(Hegde & Rokseth, 2020)

2.5.2 Application of Machine Learning on Human Factors

The application of ML to human factors focuses primarily on identifying aspects related to human inner processes or external behaviors. The rapid development and widespread application of machine learning algorithms (Ayodele, 2010) have introduced new solutions to these challenges. For instance, various issues, including sleepiness assessment (Watling et al., 2021), fatigue detection (Antwi-Afari et al., 2023; Nasirzadeh

et al., 2020), stress recognition (Jebelli et al., 2018), activity recognition (Bhat et al., 2018; Dovgalecs et al., 2010; Ma & Ghasemzadeh, 2019), personality traits detection (Evin et al., 2022) and decision-making (Evin et al., 2022), have all employed different ML algorithms to address them. Notably, the use of ML in assessing perceived risk was also observed (B. G. Lee et al., 2021). Several typical applications of ML in human factors are listed in Table 2.3.

Table 2.3: Description of published papers using machine learning to analyze human factors

Citation	Purpose(s)	Subjects	Metrics	ML	Model of better performance
(Liang et al., 2007)	Driver Cognitive Distraction	10	eye movements and performance data.	SVM logistic regression models	SVM: accuracy 0.961
(Lethaus et al., 2013)	Driver intent prediction	10	gaze data	Artificial Neural Networks (ANNs), Bayesian Networks (BNs), Naive Bayes Classifiers (NBCs)	Bayesian Networks (BNs) (according to time window length)
(Momeni et al., 2019)	Cognitive Workload Classification	24	respiration, ECG, PPG, ST	K-NN SVM DT	eXtreme Gradient Boosting (XGB):0.86
(Kumtepe et al., 2016)	Driver aggressiveness detection	83 driving sessions	visual and sensor features	SVM	SVM :0.822

Table 2.3, continued

Citation	Purpose(s)	Subjects	Metrics	ML	Model of better performance
(Li et al., 2020)	Driver Mental fatigue(multi-level)	6	eye-tracking	Support Vector Machine (SVM), Decision tree, Boosted tree, KNN LDA	SVM (accuracy: 0.795-0.85)
(Evin et al., 2022)	Personality trait	69	EDA eye-tracking data Behavioral data ECG	Decision tree (C45) Naïve Bayes (NB) Random Forest (RF) Support Vector Machine (SVM) The Neural Networks (NN)	RF
(Xiong et al., 2023)	Separation errors for air traffic controllers	10	ECG EEG	Encoder-decoder LSTM network, Convolutional Neural Network (CNN), Gated Recurrent Unit (GRU), and classic LSTM	Encoder-decoder LSTM network:0.93

Table 2.3, continued

Citation	Purpose(s)	Subjects	Metrics	ML	Model of better performance
(Fan et al., 2018)	Driving behavior (multi-classification)	50	Eye tracking	Convolutional Neural Network (CNN) Long Short Term Memory (LSTM) network, SVM Decision Tree, Random Forest	LSTM:0.87
(B. G. Lee et al., 2021)	Perceived risk classification (low and high risk level)-construction	8	EDA, PPG and ST	Gaussian SVM, K-Nearest Neighbors, Decision Tree, Bagging Tree	GSVM Low :F1 score, 0.819 High: F1 score, 0.805 Accuracy = 0.812
(Christopher & alias Balamurugan, 2014)	Predicting warning level	Not mentioned	aircraft dataset	DT, KNN, SVM, NN and NB	DT: accuracy:0.9768
(F. Li et al., 2022)	Vigilance levels assessment	14	Eye tracking	A shallow artificial neural network (SANN) SANN Linear regression Decision tree Support vector machine Bagged trees	A shallow artificial neural: 0.792(MSE)

Table 2.3, continued

Citation	Purpose(s)	Subjects	Metrics	ML	Model of better performance
(Goh et al., 2018)	evaluate relative importance of different cognitive factors	80	unsafe behavior checklists	Decision Tree (DT) Artificial Neural Network (ANN) Random Forest DT (RF) K-Nearest Neighbor (KNN) Support Vector Machine (SVM) Logistic (Regression) Naïve Bayes (NB)	DT: AUC0.976
(Ding et al., 2016)	Predicting driver's stop-or-run behavior	Not mentioned	high-resolution loop detector data	gradient boosting logit model (GBLM) statistical logit model	GBLM (0.889,0.952, 0.901 for different situations)
(Xinran Zhang & Yan, 2023)	Predicting collision (binary classification)-driving	32	EEG metrics, only driving behavior and combined EEG metrics with driving behavior	Multi-layer perceptron (MLP), Logistic regression (LR) Random forest (RF)	single time point:0.729 multi-time point: LR(0.880)
(Jebelli et al., 2018)	Stress recognition (high and low stress)	Not mentioned	EEG	SVM	SVM:0.711
(Aiello et al., 2022)	workers' movements	Not mentioned	The vibration signals	K-Nearest Neighbor (KNN)	Accuracy:0.97-0.98 Precision:0.94-0.96 F-score:0.97-0.98

Table 2.3 shows that the most used algorithms in human factors research were general machine learning, such as SVM, RF, and others, with relatively fewer instances of deep learning applications. Moreover, single or combined variables were frequently utilized as algorithm input vectors. These variables often consisted of combinations such as multiple physiological signals or integrating physiological data with survey questionnaires and behavioral data. This practice was common for both classification and regression tasks.

For example, hierarchical clustering analysis, an unsupervised machine-learning technique, was employed to evaluate operator fitness-for-duty (FFD) (Choi & Seong, 2020). This study used eye movement data and subjective fatigue ratings to perform clustering tasks. In another study, the combination of EDA, PPG and ST achieved higher accuracy (81.2%) than only using EDA or PPG for distinguishing between low and high levels of perceived risk. Moreover, SVM with Gaussian kernel function performed better than DT, KNN and Bagging Tree (B. G. Lee et al., 2021). However, the integrated metrics only sometimes obtained better results. For example, EEG and behavioral variables were utilized to identify unsafe drivers. Multilayer Neural Networks (MLP), one representing the algorithm of ANN, reached better prediction effectiveness compared with Logistic regression (LR) and Random Forest (RF) (Zhang & Yan, 2023). However, a single EEG performed better than a combination of EEG and behavior variables.

There is a trend in utilizing multi-physiological signals to address various issues in human factors (Awais et al., 2017; Fan et al., 2020; Zhou et al., 2018). This aimed to tackle facets of the problem, compensate for individual method limitations, and ultimately achieve enhanced performance (Choi & Seong, 2020).

2.5.3 Application of Machine Learning on Perceived Risk

The statistical analysis of physiological response data has significantly enhanced our understanding of workers' risk perception. For instance, ANOVA is used to detect significant differences in EDL and EDR between high and low-risk situations in

construction (B. Choi et al., 2019) and to analyze perceived risk tolerance under different conditions in construction, such as no, slight and dense smoke (Fu et al., 2021), as well as during various accident stages, including control, precursor, and outcome periods (Barnard & Chapman, 2016). Nevertheless, statistical analysis is inadequate to explore the exceedingly complex nonlinear relationship between physiological changes and perceived risk. The advancements in machine learning and their successful application in the field of human factors provide exciting potential for identifying perceived risk.

2.5.3.1 The application of ML for assessing perceived risk in social science

In the context of risk perception, ML techniques have the potential to identify perceived risk in social science. Machine learning techniques, particularly deep learning, have been employed to assess public risk perception. Social media platforms provide a rich source of data on how people perceive risks, as individuals often discuss and share their thoughts and feelings about various risks on social media (Guan & Chen, 2014; Gui et al., 2017; Wu & Cui, 2018). In addition, the word distribution structure in natural language data was utilized to discover rich representations of numerous naturalistic risk sources (Bhatia, 2019). However, the input vectors in these studies primarily focused on external factors, such as website search and social information, rather than physiological signals (Lawless, 2022; Tang et al., 2022; Weber, 2017; Xie & Xue, 2022).

2.5.3.2 The application of ML for assessing perceived safety risk

In terms of safety risk perception, the four most related research have shown their contribution to the application of ML in assessing risk perception. One focused on the drivers' perceived risk (Ding, Ghazilla, et al., 2022; Ding, Raja Ghazilla, et al., 2022; Ping et al., 2018), while the other three examined the perceived risk among construction workers (B. Choi et al., 2019; Byungjoo Choi, Gaang Lee, et al., 2019; Jeon & Cai, 2021; B. G. Lee et al., 2021). The findings of these studies have instilled significant confidence

in the development of a fully data-driven perceived risk assessment model. Therefore, the following section will provide a detailed analysis of these findings.

(a) The lack of emotional responses as input vectors

In terms of input vectors, emotional responses are rarely considered. The most used indicators were personal operational data and physiological signals. For instance, the study incorporated vehicle parameters, such as steering angle, brake pedal position, gas pedal position, longitudinal acceleration, and presence of parked vehicles, pedestrians, and bicycles, as well as road characteristics (including curves, uncontrolled and controlled intersections) (Ping et al., 2018). These data were collected during the subjects' real driving experiences and treated as time series data to represent the process of risk perception. However, for construction workers' subjective risk assessment, wearable sensors were employed to collect workers' physiological signals during their regular tasks in the field. One exclusively used Electrodermal Activity (EDA) (Byungjoo Choi, Gaang Lee, et al., 2019), or EEG (Jeon & Cai, 2021) as the input vectors, while the other employed a combination of physiological signals, including EDA, Photoplethysmography (PPG), and Skin Temperature (ST) (B. G. Lee et al., 2021). The same research team conducted the two studies. Unfortunately, emotional indicators were not included as input vectors to classify perceived risk since induced emotion during human perceiving risk had significantly influenced the assessment process and results. It can be seen from the separation of System 1 and System 2.

(b) The employment of different ML algorithms

Regarding algorithms, basic ML algorithms have exhibited exciting results. For drivers' risk perception (Ping et al., 2018), the performance of a long short-term memory-based method (LSTM) demonstrated superior performance compared to other algorithms like Neural Networks (NN) and Support Vector Machines (SVM) due to its ability to effectively handle time series data. The two studies focusing on construction workers

evaluated the different algorithms by classification index, such as accuracy, precision and F1 score. Among these algorithms, K-Nearest Neighbors (KNN) showed the best performance, surpassing Decision Trees (DT), Logistic Regression (LR), Gaussian Support Vector Machine (GSVM), Subspace KNN (SKNN), and Bagging Tree (BT). KNN achieved an accuracy of 76.9% in classifying high and low-risk levels using electrodermal activity (EDA) as input indicators (B. Choi et al., 2019). Another study demonstrated that the Gaussian SVM was a better classifier for classifying high- and low-level risk by employing the combination of physiological indicators, compared with KNN, DT, and BT (B. G. Lee et al., 2021). So, it can be inferred that the basic machine learning models can effectively capture the relationship between physiological data and risk situations with significantly different levels of risk.

(c) The deficiency in classifying situations with close risk degrees

ML models for classifying risk situations with closely matched risk degrees still require future research, even though they have demonstrated better performance in predicting high and low-risk levels. In a study on drivers' risk perception (Ping et al., 2018), risk levels ranged from no feeling of risk to an extreme sense of risk, as ranked by participants while watching videos. While the study considered three scenarios beyond significant risk disparities: slight, moderate, and intense feelings of risk, the input data did not include physiological signals, preventing an evaluation of their predictive ability. In the case of construction workers' risk perception (B. Choi et al., 2019; Byungjoo Choi, Gaang Lee, et al., 2019; B. G. Lee et al., 2021), typical binary classification tasks were performed, focusing solely on high and low-risk levels. They were determined by the authors based on video-recorded workers' activities. The three research primarily utilized EDA or the combination of three physiological signals to classify risk levels. However, they did not consider risk degrees that were closely matched. Only when physiological responses can accurately identify risk situations with minimal differences in risk degree

can the significant potential of machine learning models in subjective risk assessment be fully applied.

2.5.4 Discussion of Machine Learning Application

Based on the above content, it can be concluded that people experience physiological and emotional changes regulated by the autonomic nervous system in risky situations. These changes are dynamic, multidimensional, and highly complex. Establishing a model to recognize the relationship between these changes and perceived risk levels is not a simple task. However, machine learning offers several advantages over traditional statistical analysis methods. For example, it requires fewer assumptions about the data distribution. It can handle multidimensional variables or wide data (where the number of input variables exceeds the number of samples) and has strong nonlinear fitting capabilities. Clearly, these advantages highlight the feasibility of machine learning in assessing perceived risk. The main points were discussed regarding the consensus on the application of machine learning in risk assessment.

Broad applicability: Machine learning has been applied in various stages of risk assessment, including risk identification, risk analysis, and risk evaluation. This widespread utilization indicates a consensus that machine learning can effectively address various aspects of risk assessment.

Performance and validation: Certain machine learning algorithms, such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Decision Trees (DT), and Random Forest (RF), have been consistently demonstrated to deliver satisfactory performance in risk assessment tasks. The validation of these algorithms across multiple studies strengthens the consensus on their effectiveness.

Exploration of human factors: The specific research on human factors using machine learning differs from the traditional approach of utilizing historical data on human operations. Instead, it utilizes diverse data sources such as real-time

multidimensional physiological indicators, the fusion of different physiological indicators, behavioral measurements, and environmental parameters, even complex images or real-time videos. Machine learning goals vary from predicting or classifying cognitive biases, fatigue, alertness, and stress to certain unsafe behaviors. It indicates a consensus on the potential of machine learning to provide valuable contributions to understanding and managing human factors in risk assessment.

Opportunities for perceived risk: There is limited research specifically focused on the application of machine learning in subjective risk perception assessment. Nevertheless, some studies have started exploring data-driven assessment models based on human physiological and psychological responses under different risk levels. Machine learning techniques provide an excellent opportunity for this inherently subjective and implicit risk perception assessment. However, machine learning algorithms must continue to perform well only when there is a minimal disparity in the magnitude of perceived risk, thus ensuring the prospective application of machine learning.

2.5.5 The Comparison of ML with Other Methods

It can be concluded that questionnaire originated in social science and engineering risk assessment started with safety science were methods commonly used to evaluate risk. ML, with the rapid development of AI and data accumulation, had been a new way to assess risk and shown great power. ML would be used in this research to assess risk. To display its suitability, the comparison of ML with other methods used in risk perception assessment was listed in Table 2.4.

Table 2.4: The comparison of ML with other methods used in risk perception

No.	Questionnaire	Engineering risk assessment	ML
Application areas	Social science Engineering areas	Engineering areas	Social science Engineering areas
Advantages	It was simple to collect data. The data collected was subjective and biased. It was easy to be conducted.	The assessed results were often regarded as objective and can reflect the real level of risk.	It required fewer assumptions about data distribution. It can handle multidimensional variables or wide data. It had strong nonlinear fitting capabilities. The results of the assessment were close to the subjective perceived risk.
Drawbacks	The assessment results were often considered with subjectivity and bias.	Its application needed specialized engineering. It was hard to evaluate the subjective and perceived risk which generated in humans' brain, not existed in machines or environments.	The collection of part data was difficult and may cost more time. The progress of model training may be long, especially for big data. The quality of data had impact on model performance.

2.6 Individual Differences of Risk Perception

The variability in individuals' risk perception of a given hazardous situation arises from its inherently subjective nature and the impact of influencing factors. As physiological signals can reflect the process of risk perception, individual differences can be reflected in the physiological signals accompanying the process of risk perception. However, there is limited research on this specific issue. Encouragingly, recent studies have begun exploring the potential individual differences that may exist in risk perception. These studies span multiple disciplinary areas, with safety, human factors, and ergonomics being the most prominent among them (Bagley et al., 2023).

Firstly, various research has highlighted the personal differences in risk perception, which is a highly personal process based on an individual's understanding of risk (Brown, 2014). These differences were observed in some studies. For instance, research on pilots showed that those with better safety performance exhibited more stable inter-beat intervals (IBI) and higher skin conductivity levels (SCL) during hazardous scenarios (Gao & Wang, 2020). Another study explored participants' route choices in different levels of stimulated smoke using immersive virtual reality (VR) and found that individuals with a high tolerance were more inclined to take a dangerous shortcut. However, an individual's attitude towards risk can be dynamic, and their everyday risk preference in a low-risk context may not necessarily align with their decision-making in high-risk situations (Fu et al., 2021). While these studies described individual differences, they did not offer a scientific perspective.

Significantly, one research focused on resolving this issue (B. Choi et al., 2019). This research used skin conductance to assess construction workers' risk perception. It observed differences in skin conductance responses between high and low risk levels and determined that these differences were partially attributed to distinct individuals. Hierarchical Linear Modeling (HLM) was employed to determine its existence. Therefore, there is preliminary experimental evidence indicating the existence of personal differences in risk perception. Nevertheless, further research is needed to confirm these findings.

The existence of individual differences in risk perception during risk assessment is of paramount importance, as it has been demonstrated that significant variations in physiological responses associated with the risk perception process exist. However, further investigation is required to ascertain whether the origin of these differences lies in individual variations or different risk situations.

2.7 Gap Summary

Figure 2.6 organizes the principal findings and limitations identified in the literature review to present the research problems related to risk perception.

Risk perception originated in the social sciences, but its direct impact on decision-making and behavior has rendered it a significant research issue within the field of safety science. Risk perception and engineering risk assessment methods represent two fundamentally distinct paradigms for quantifying risk. In particular, safety behaviors mainly rely on emotional perception rather than rational calculations of risk. Consequently, there is a pressing need for new assessment paradigms for human-perceived risk. The rapid advancements in wearable physiological and emotional measurement technologies within the field of ergonomics provide a promising opportunity for advancing research on risk perception. So, the primary objective of this research is to develop machine learning models that predict perceived risk level using a combination of physiological and emotional indicators as input data. These indicators can be obtained in different risk and experimental scenarios. According to the main findings of the literature review, the following research gaps and their corresponding issues were summarized.

Firstly, it is widely recognized that specific physiological signals display significant differences between high and low risk situations. However, whether these observed differences hold statistical significance in situations with only minor risk variations remains uncertain, underscoring the necessity for further investigation. Additionally, there is a need to study the representation, and evolving patterns of emotions induced by risk situations. Meanwhile, exploring the more comprehensive and abundant neuromodulation mechanisms and possible individual differences associated with risk perception is warranted.

Secondly, progress has been made in establishing machine learning models to classify

high and low-risk scenarios based on physiological responses. However, further research is required to determine whether such models can still perform well in classifying risk situations with modest differences in their degree of riskiness. It will further substantiate the successful utility of machine learning applications. In addition, it has been demonstrated that a combination of multidimensional physiological indicators can differentiate risk situations. However, there is limited research regarding which specific indicators are more sensitive for classifying risk within the classification process.

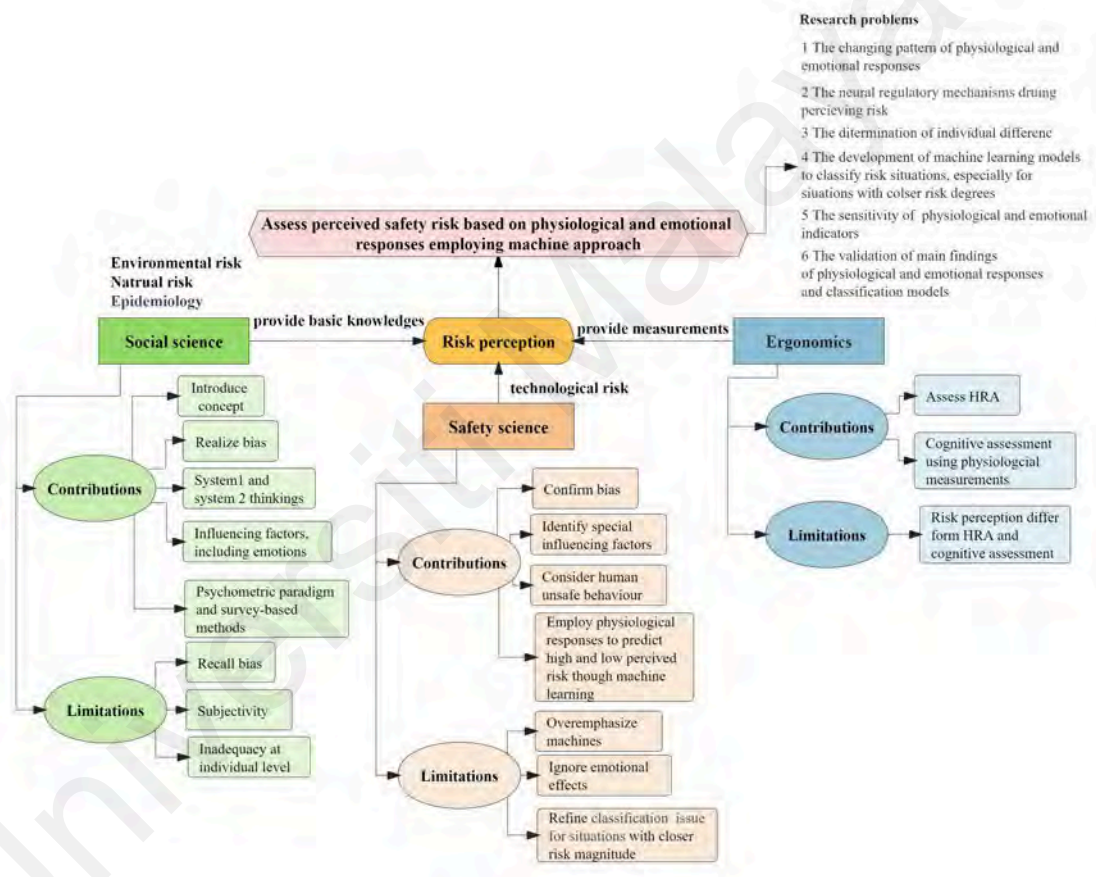


Figure 2.6: Formulation of research problems based on the literature review

Thirdly, it is essential to validate the performance of the proposed machine learning model for classifying risk perception across diverse risk scenarios, as this confirmation can demonstrate its generalization capability. Moreover, this validation will establish the stability of the inherent relationship between physiological, emotional responses and risk levels, confirming that the model's classification capability remains consistent across different scenarios and individuals.

CHAPTER 3: METHODOLOGIES

3.1 Introduction

This chapter is to design experiments that collect human response data in various risk situations, focusing on situations with closer risk degrees. The field of driving has consistently been a popular choice for studying risk perception. Consequently, the initial experiment was conducted within a driving scenario. The subsequent validated experiment, explicitly simulating a construction site, was undertaken to confirm the robustness of findings from the initial experiment.

Furthermore, this chapter elucidates statistical analysis methods used to analyze the experimental data. Additionally, it elaborates upon the standard procedures of machine learning algorithms.

3.2 The Comprehensive Research Design

Based on the literature review, six research problems were identified and depicted in Figure 2.6. Correspondingly, three primary objectives were formulated in alignment with the identified research problems. Figure 3.1 depicts a comprehensive flow chart for achieving these goals. The initial and validated experimental scenarios, limited to driving and construction sites, shared similar their design processes. Subsequent steps were conducted to acquire physiological and emotional data. These steps included screening appropriate stimuli, recruiting subjects, determining physiological measurements, designing the experiment process and selecting physiological sensors, which were conducted to acquire physiological and emotional data.

The first and second objectives were achieved through the analysis of data from the driving experiment. These findings were validated by analyzing the experimental data obtained after changing to the construction experiment scenario. In this way, the third objective was also accomplished.

A detailed explanation of the research design will be provided in the following section of this chapter.

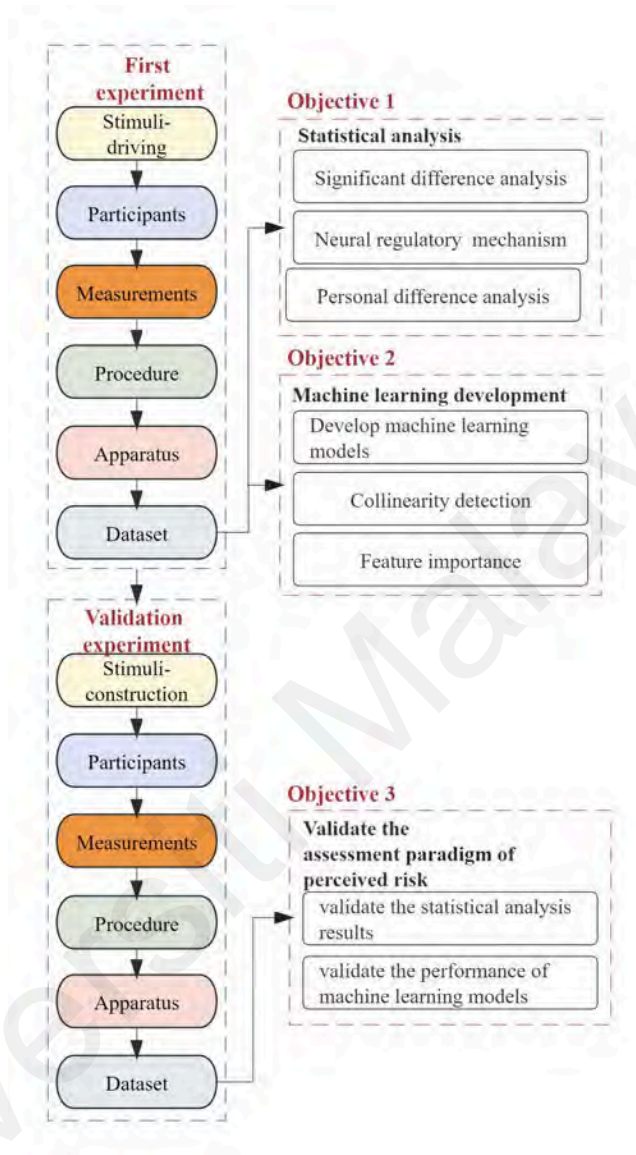


Figure 3.1: The comprehensive research flow chart

3.3 Stimulus





There was no doubt that measuring physiological and psychological signals in risk situations was challenging to perform in real-world settings (Brookhuis & de Waard, 2010). A laboratory setting can provide a choice. Videos were often considered adequate inducing specific situations (DASDEMİR et al., 2017). Besides, virtual reality (VR) was often regarded as a more effective way to simulate real scenarios (Perlman et al., 2014). However, in Australia, risk perception measurement for individuals applying for a driver's

license was conducted through driving videos (Hill et al., 2019). So, videos, as stimuli causing a physiological or psychological response (Minhad et al., 2017), were also utilized in this research.

Seven driving videos were downloaded from the internet, which portrayed risky situations from the driver's perspective. Three criteria were applied to select suitable videos. Firstly, the clips depicted the perceived likelihood of an accident occurring (probability judgments) and its consequences (severity judgments) (Christian et al., 2009). In other words, the entire process of presenting the danger, the occurrence of the accident, and its consequences should be comprehensively included. Secondly, each clip only depicts a single key event (Iqbal et al., 2016), which help subjects pay attention to specific risk situation. Thirdly, the duration of the videos should be similar, considering the effect of time. The duration of the seven videos was approximately two and a half minutes, with differences of less than 10 seconds. Subsequently, 100 college students with driving experience were recruited during the preparation stage of the experiment to assess the effectiveness of situational induction. Participants used a scale from 1 to 10 to rate the authenticity of each video based on their own driving experiences: the higher the score, the closer the video resembled real driving situations.

Since risk perception was considered a spectrum rather than a binary concept, it fluctuated across various risk levels, from no risk to significantly high-risk levels (Sitkin & Pablo, 1992). Additionally, similar video clips' content has been removed. Finally, four videos were ultimately selected. One of the clips depicted no unexpected incidents and was used as a baseline for comparison with the others. The details and some screenshots representing essential events of the selected videos are described in Table 3.1.

Table 3.1: A simple description of four driving clips

No.	Description
1	<p>The driver was leisurely driving and saw teachers and children passing through the zebra crossing. He stopped ahead of time until they finished crossing. There was no abnormal situation.</p> 
2	<p>While driving, the driver failed to perceive the oncoming cyclists ahead of time. He had to brake hastily to avoid hitting them. Thankfully, the car stopped in time, and no one was injured. However, the driver seemed scared.</p> 
3	<p>During rush hour, the driver almost collided with the car in front. Fortunately, he braked sharply in time to avoid a rear-end collision. However, he still appeared frightened and took a long breath.</p> 
4	<p>Suddenly, a car turning a corner crashed into another car coming from the side direction. The front windshield of the car was shattered, and glass shards were scattered. Unfortunately, the driver was bleeding and the car window was stained with blood. Obviously, there were severe consequences.</p> 

The risk levels of the selected clips needed to be ranked. 123 students with driving experience were recruited to give the risk scores for each clip on a scale from 1 to 5, with higher numbers indicating an increase in risk. Two clips had similar risk rankings, scoring 2.45 and 2.85, respectively. Despite the different situations depicted in the clips, the risk scores assigned by the students were very close, indicating similar risk levels. The first clip was considered less hazardous, scoring 1.87, and served as a suitable baseline measurement. Conversely, the last clip, scoring 3.97, was the most dangerous among the

clips. However, it still fell within the students' acceptable range, and no apparent discomfort was observed.

3.4 Participants

In related research, eight workers were recruited to participate in experiment (Byungjoo Choi, Houtan Jebelli, et al., 2019; G. Lee et al., 2021). Their physiological data were collected during high and low risk situations. The number of data used to train ML models were 232 and 224, respectively. Considering there were four clips, 60 subjects are suitable to record their response data. This will result in 240 samples for training models. The number of samples is similar to that in related research.

Regarding the selected standards, previous studies have demonstrated that risk perception in driving situations was influenced by driving experience and became more sensitive over time (Tanida et al., 2018). Consequently, we recruited 60 student volunteers with three years of driving experience, ensuring that experience-related implications were addressed. The age of participants ranged from 21 to 23 (mean: 22.4; SD: 1.35), with ten females and fifty males. Notably, the subjects had not seen the clips yet at the selecting stage. The order of their participation during the experimental measurement periods was randomized, and all subjects had normal vision or corrected-to-normal vision. Prior to the commencement of the experiment, participants were instructed that situations exceeding their mental capacity were strictly prohibited. All participants provided informed consent and engaged in practice trials to familiarize themselves with the experimental procedures and the presentation of stimuli. Furthermore, they were required to have a good night's sleep before the testing session.

3.5 Measurements

3.5.1 Physiological Measurements

Based on fundamental considerations for physiological measurement, it is believed that combining various physiological signals would be more effective in identifying risk

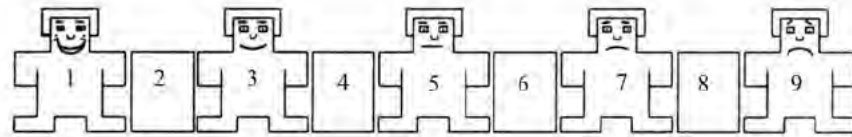
(B. G. Lee et al., 2021). Hence, multiple physiological measurements were employed in the present research. According to the previous literature review, electrocardiogram (ECG) (Panicker & Gayathri, 2019), electrodermal activity (EDA) (Kreibig, 2010), and pupillary response (Gajardo et al., 2019) were popular methods for measuring risk and were utilized in this study. Additionally, skin temperature (ST) was introduced at the construction site to assess perceived risk (B. G. Lee et al., 2021), and it was included as well. It was important to emphasize that these physiological signals were obtained while subjects were watching video clips, and the data were synchronized with the timeline of the clips.

3.5.2 Emotion Measurements

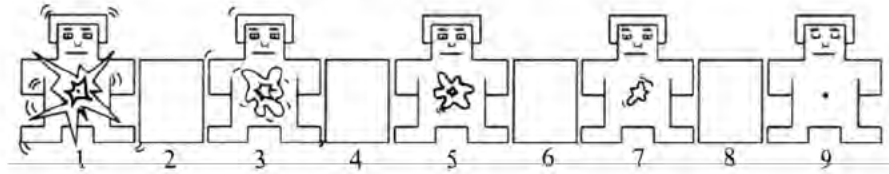
The PAD emotion model may be a viable choice due to the complication and fluctuation of emotional responses in risky situations. It assigns numerical values ranging from 1 to 9 to represent the continuous emotions across three dimensions: pleasure, arousal, and dominance. This model maps emotion onto three distinct spaces, allowing for the quantification of subjective emotions and facilitating comparisons between different situations. Moreover, this approach to labeling emotions helps circumvent the difficulties in dissecting highly fused emotions.

Combining the three dimensions of pleasure, arousal, and dominance gives a more comprehensive understanding of how individuals experience emotions in response to various risk scenarios. This approach provides a holistic perspective on the emotional experiences of individuals in different risk-related situations.

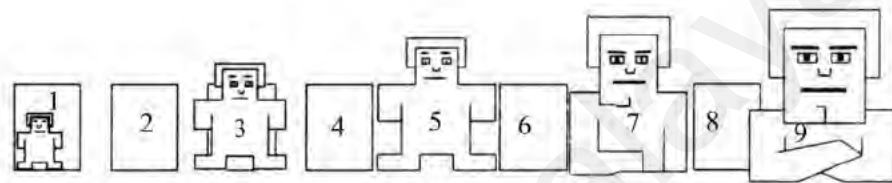
The process of scoring emotional experiences is easy to follow. Participants assign a score to represent their real feelings based on the Self-Assessment Manikin (SAM) images. The three Self-Assessment Manikin (SAM) images are shown in Figure 3.2.



a) the dimension of pleasure



b) the dimension of arousal



c) the dimension of dominance

Figure 3.2: The Self-Assessment Manikin (SAM) visuals

The meanings of the three dimensions (P, A, and D) have been introduced in the literature review section. To efficiently enable participants to give scores for the three dimensions of PAD based on their real emotional experiences within a short period, the following points were emphasized during the introduction.

Firstly, participants were instructed to start by focusing on the central figure representing the intermediate state for each dimension. If their emotional feeling was more substantial than the figure, they should move towards the right; if it was weaker than the figure, they should move towards the left. Secondly, if participants' feelings fell between two figures for any dimension, they were encouraged to assign a score corresponding to the value between them. Thirdly, it was acknowledged that understanding the connotations of the D dimension might be more challenging than the Pleasure and Arousal dimensions. Therefore, a more detailed explanation was provided to ensure the effectiveness of the proposed PAD model.

These instructions and explanations aimed to facilitate participants in accurately scoring the three dimensions based on their personal emotional experiences, ensuring the data's reliability and validity.

3.6 Procedure

Firstly, the experimental procedures were explained to the participants, including the stimulus materials used and the identification of the participants after the survey period. Detailed instructions and illustrations were provided on how to rate the P (Pleasure), A (Arousal), and D (Dominance) scores.

Secondly, the participants were informed that they would be watching four driving clips from the driver's point of view. They were instructed to imagine themselves as the driver and to remain seated still throughout the entire experiment. The clips were displayed on monitors, and the participants were asked to observe each clip.

Thirdly, the participants were equipped with physiological sensors and were seated in a comfortable position. They were given practice trials to familiarize themselves with the experimental process and stimulus display.

Fourth, once all the preparations were completed, the formal experiment commenced. Each participant was given five minutes of relaxation before watching the four video clips from the driver's perspective. They were instructed to watch the clips as if they were the driver and to remain still throughout the experiment. It was important to note that the baseline clip was shown after the relaxation segment, and the playback order of the remaining three clips was randomized. A one-minute interval was introduced between each video clip to ensure that the previous clip did not influence participants' physiological state. The order of stimulus exposure was randomly determined for each participant.

Finally, at the end of the experiment, a five-minute relaxation period was allocated to each participant. Following the experiment, the apparatus used for physiological

measurement was removed, and PAD values were given. It completed one trial and repeated the process for each subsequent participant. The overall process is illustrated in Figure 3.3.

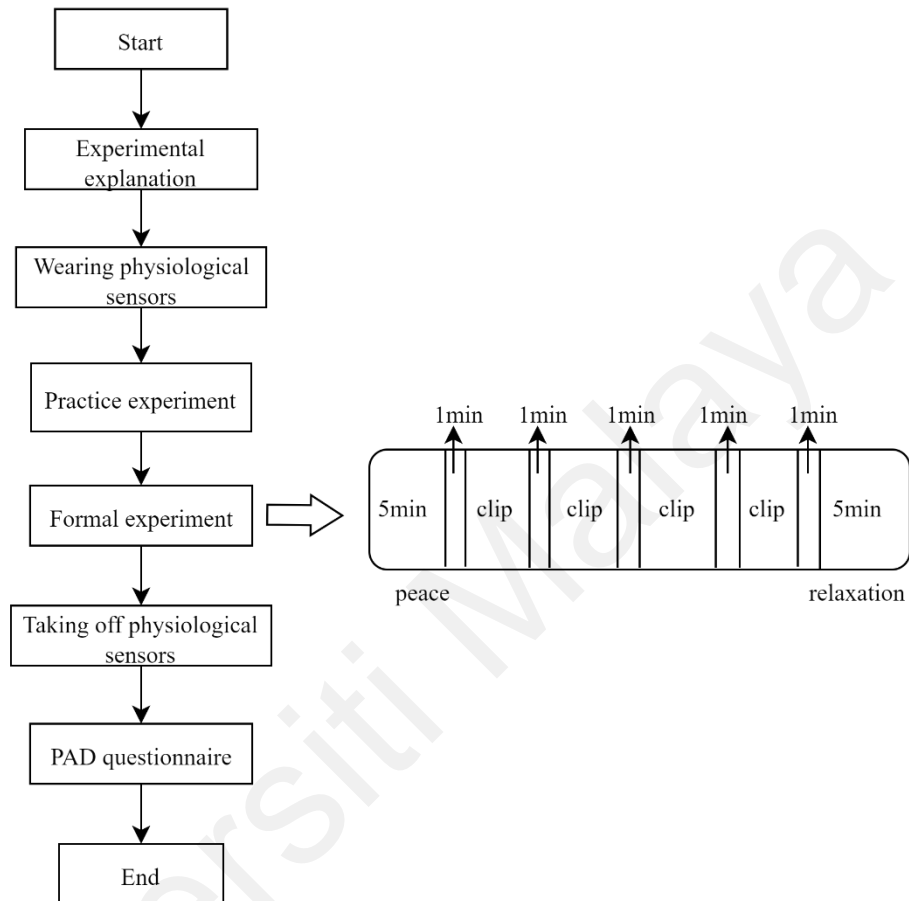


Figure 3.3: Diagram of experimental procedure

3.7 Apparatus

Experimental equipment is essential for displaying stimuli and simultaneously measuring multiple bio-signals to record data and analyze the effects of the stimulus on the subjects. The ErgoLAB ergonomic equipment is suitable for fulfilling this requirement as it enables synchronous transfer and recording of multi-parameter physiological data through connected USB devices in real time. The measuring system consists of two main components: software and hardware.

3.7.1 Software of ErgoLAB

Designing the experiment procedure and entering basic information about the participants can be accomplished during the preparation stage using the provided software. The software enables the presentation of stimuli, acquisition of data from physiological sensors, and recording of the obtained data during the experiment. Once the data collection phase is completed, the software also offers functionalities for data preprocessing, calculation of physiological indices, and data export.

Due to their unique characteristics, it is important to note that different physiological signals require specific preprocessing methods. The software provides predefined preprocessing techniques tailored to each physiological signal. These methods are outlined in the tables below, which offer a comprehensive guide for preprocessing and extracting indices based on the characteristics of the physiological signals.

The preprocessing methods of EDA and specific values are shown in Table 3.2.

Table 3.2: Parameters of EDA signal preprocessing methods

No.	Method	Parameters	Values
1	Filter: Gauss	window size/ms	5
2	Low pass	Cut off /Hz	5
3	SCR extraction	Peak detection sensitivity	Middle
		Maximum rise time/s	4
		Maximum half decay time/s	4
		Maximum amplitude/ μ s SCR	0.03
		Window size/ms	100

The preprocessing methods of ECG and corresponding values are described in Table

3.3.

Table 3.3: Parameters of ECG signal preprocessing methods

No.	Method	Parameters	Values
1	Filter: Wavelet denoise	Intensity	medium
2	Amplitude normalization	Intensity	middle
3	High pass	Cut off /Hz	1
4	Low pass	Cut off /Hz	100
5	Band stop	Cut off /Hz	50
6	R-peak extraction	Maximum heart rate/bpm	120
		R-peak mark threshold/bpm%	70
7	Ectopic detection	Median/samples	4
8	Ectopic correction	Method	median
		Window size/samples	5

Eye-tracking data preprocessing methods and parameter settings are listed in Table

3.4.

Table 3.4: Parameters of eye tracking raw data preprocessing methods

No.	Method	Parameters	Values
1	Noise reduction	Method	Moving Median filter
		Window size/ms	3
2	Discard short fixations	Minimum fixation duration/ms	60
3	Pupils process	Method	Linear
		Minimum pupil diameter /mm	2

By utilizing the software, researchers can efficiently handle the entire experimental process, from designing the procedure and collecting data to preprocessing and analyzing the acquired physiological signals. This integrated approach streamlines the data processing workflow and ensures the accuracy and reliability of the results.

3.7.2 Hardware of ErgoLAB

The system's hardware components consist of a computer host and a Light Emitting Diode (LED) monitor, as depicted in Figure 3.4. The computer host is equipped with the

software mentioned above system installed. The LED monitor is 18 inches long and is used for displaying stimuli during the experiment.



Figure 3.4: The image of computer host and LED monitor

Additionally, various physiological sensors are used to measure different variables, as shown in Figure 3.5. Numerous sensor straps helping to fix the sensor and the charger are displayed.



Figure 3.5: The different physiological measurement sensors

To measure ECG, three electrodes are attached to the participants' chest (showed in Figure 3.6).

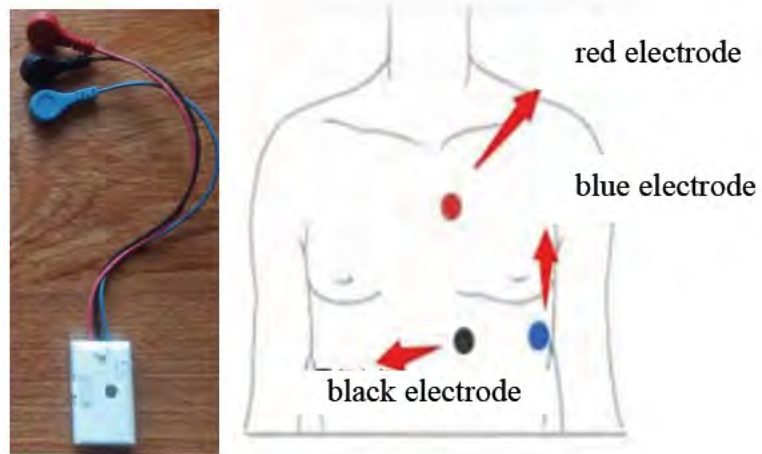


Figure 3.6: The wireless ECG sensor and its wearing methods

EDA is measured using two electrode sheets attached to the index and middle fingers of the non-dominant hand and is secured with an adhesive strap (shown in Figure 3.7).



Figure 3.7: The wireless EDA sensor and its wearing methods

PD is recorded using an eye-tracking device, which measures the participants' eye movements while looking at the screen (shown in Figure 3.8).



Figure 3.8: The Screen-based eye-tracking device

ST is measured using a thermistor sensor attached to the participant's index finger of the right hand, which is also secured with an adhesive strap (shown in Figure 3.9).



Figure 3.9: The wireless ST sensor and its wearing methods

The system combines a computer host, LED monitor, and various physiological sensors. This integration enables the system to deliver stimuli and accurately capture and record physiological responses during the experiment. The hardware components ensure smooth operation and reliable data acquisition. Importantly, physiological signals can be continuously recorded during video presentation and transmitted synchronously using wireless. A device connected to the host via a USB interface facilitates wireless data transmission.

3.7.3 Physiological indices

Because the level of perceived risk tends to be sustained for a specific duration and changes slower than stress and physical exertion (B. G. Lee et al., 2021), the recorded physiological data during each clip using synchronous acquisition techniques serve as an analyzing segment that reflects the averaged physiological reactions for each risk situation. Subsequently, the four segmental physiological signals were processed using filtering, noise reduction, and physiological index extraction parameters. Physiological indicators in different scenarios can be exported, with the meanings and physiological significance of each index listed in Table 2.1.

After calculation, there are a total of 17 indicators, including two EDA indicators (EDL and EDR), four ECG temporal indicators (HR, SDNN, RMSSD and SDDSD), five ECG frequency indicators (ULF, VLF, LF, HF and LF/HF), four ECG nonlinear indicators (SD1, SD2, A++ and B--), one skin temperature (ST), and one pupil diameter (PD).

3.8 Validation Experiment

A similar experiment was designed to confirm the analysis results, maintaining the same overall paradigm while shifting the risky situations to construction sites. Therefore, this section will primarily elaborate on the stimulus presented in the video clips. The video clips were downloaded and selected using the same criteria as before, and their risk levels were rated on a scale from 1 to 5. The clips were presented in an animated format. Four clips were chosen as stimuli, each for approximately three and a half minutes. Similar to the previous experiment, one clip was designated as a baseline measurement.

To ensure the acceptability of the video plots, a survey questionnaire was conducted to gauge participants' reactions. The responses indicated that the video plots were within an acceptable range and did not elicit significantly uncomfortable feelings. The descriptions of the video plots are provided in Table 3.5. The first scene, representing a relatively benign scenario, served as the baseline with the lowest risk rank of 1.83. The fourth scene had the highest risk score at 3.85. The risk values for the two middle scenes were close, with scores of 2.55 and 2.69, respectively. Comparing the risk values, the risk levels of the three videos portraying negative consequences in the construction scenario were relatively similar in contrast to the driving scenarios.

Table 3.5: A simple description of four construction clips








No.	Description
1	<p>This clip mainly introduces the usage method of safety harnesses. It is the most used personal protective equipment in construction sites. The video covers its appearance check, usage scenarios, method, and precautions.</p> <div style="display: flex; justify-content: space-around;">   </div>
2	<p>It depicts a construction site safety officer performing a safety inspection and discovering that the scaffolding was not constructed according to the relevant technical requirements. The workers at the site stated that they would rectify the situation within two days, but in practice, they needed to reinforce it as required. As a result, the formwork collapsed, causing the entire completed section of the building to collapse, and a large amount of dust rose on the scene. The clip shows that there are no personnel injured in the incident.</p> <div style="display: flex; justify-content: space-around;">   </div>
3	<p>The video clip showed a construction worker removing the protective paper from the window frame and cleaning the frame. After taking the elevator to a high floor, he walked to the window, stood directly on the frame without wearing a safety harness, and began removing the protective paper. The worker's footing slipped during the task, causing him to fall from the high place. The entire process of the person falling and the blood on the ground after the fall were captured in the video.</p> <div style="display: flex; justify-content: space-around;">   </div>

Table 3.5, continued

No.	Description
4	<p>Three workers dismantled the scaffold poles and transported them from the upper floor to the ground via an elevator. However, they thought it was the weekend, and nobody was working at the construction site, so one worker decided to go upstairs and throw the poles down from a height while the other two workers waited on the ground to collect the thrown poles. During the operation, one of the workers sleeping on the unfinished first floor was awakened by the sound of the pole falling to the ground. He came out to see what happened but did not notice the pole falling from a height and was hit in the head, causing him to fall to the ground.</p> 

60 college students with essential construction knowledge were recruited to participate in this experiment. Their ages ranged from 21 to 23, with ten females and fifty males.

3.9 Data Analysis

3.9.1 Statistical Analysis

The nature and distribution of the data can influence the choice of statistical methods. Therefore, in this study, different statistical methods were employed to analyze discrete emotion indicators and continuous physiological indicators.

3.9.1.1 Discrete analysis of PAD emotions

The P, A, and D values range from 1 to 9, allowing only nine discrete integers. As a result, the three-dimensional emotion variables are categorical. The data for these variables in the two scenarios, including four situations, did not follow a normal distribution. Therefore, the non-parametric Kruskal-Wallis rank sum test was employed. If significant differences were observed for a variable, further analysis using the non-parametric Mann-Whitney U test was conducted to identify which two groups differed significantly.

The calculation progress of Kruskal-Wallis test is shown as follows.

Rank all the data: Combine all groups' data into a single list and assign ranks to each data point. If there are ties, assign the average rank to the tied values.

Calculate the sum of ranks for each group: Sum the ranks for the data points in each group.

Compute the H statistic: Use the following formula.

$$H = \left(\frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} \right) - 3(N + 1) \quad (3.1)$$

where:

N is the total number of observations across all groups.

k is the number of groups.

R_i is the sum of ranks for the i -th group.

n_i is the number of observations in the i -th group.

Convert the H statistic to an η^2 value: This can be done using the following formula:

$$\eta^2 = \frac{H-k+1}{n-k} \quad (3.2)$$

Convert η^2 to Cohen's f : The effect size of Cohen's f is related to η^2 , and can be done using the following equation.

$$\text{Cohen's } f = \sqrt{\frac{\eta^2}{1-\eta^2}} \quad (3.3)$$

The statistical analysis for this study was conducted using SPSSPRO, a free online platform. This choice was based on its capability to directly calculate effect sizes for pairwise comparisons (SPSSPRO, 2021).

Effect size is commonly employed to assess the magnitude of difference between variables in statistical analyses. A larger effect size indicates a more significant difference or relationship between variables, while a smaller effect size suggests a weaker association. In the case of the Kruskal-Wallis test, SPSSPRO utilizes Cohen's f value to represent the effect size, with critical points of 0.1, 0.25, and 0.40 denoting small, medium,

and large effect sizes, respectively. Similarly, Cohen's d value is used to gauge the effect size of pairwise comparisons, with critical points of 0.2, 0.5, and 0.8 indicating small, medium, and large effect sizes, respectively.

3.9.1.2 Continuous analysis of physiological responses

Since physiological indicators (dependent variables) are all numerical data, one-way (risk level) repeated-measures ANOVA is used for difference analysis. This research is conducted within subjects' analysis to investigate the differences in individual' physiological signals- EDA, ECG, PD and ST under different risk levels. Specifically, EDL and EDR are indicators derived from EDA. HRV includes time-domain indicators such as HR (Heart Rate), SDNN (Standard Deviation of NN Intervals), RMSSD (Root Mean Square of Successive Differences), SDSD (Standard Deviation of Successive Differences), frequency-domain indicators of HRV such as ULF (Ultra Low Frequency), VLF (Very Low Frequency), LF (Low Frequency), HF (High Frequency), and LF/HF ratio. Nonlinear indicators in HRV include SD1, SD2, A++, and B--. There are 17 indicators together with Eye movement indicators, including PD (Pupil Diameter) and ST (Skin Temperature).

For various physiological indicators, non-parametric testing methods are used if the data approximately follows a normal distribution and passes the Sphericity tests. If the data does not meet the above conditions, test results are adjusted using the Greenhouse-Geisser correction. Furthermore, if there are significant differences at different levels, Bonferroni post hoc tests are used.

The effect size indicating the degree of difference in one-way repeated measures ANOVA (η^2) and post hoc tests (t-test) (Cohen's d) can be used to measure the effect size of pairwise comparisons. They can be calculated by the two formulas.

$$\eta^2 = \frac{SS_{effect}}{SS_{effect} + SS_{error}} \quad (3.4)$$

SS_{effect} is the sum of squares for the effect (within-subjects factor).

SS_{effect} is the sum of squares for the error term associated with the effect.

$$\text{Cohen's } d = \frac{\bar{X}_1 - \bar{X}_2}{S_d} \quad (3.5)$$

\bar{X}_1 and \bar{X}_2 are the means of the two conditions being compared.

S_d is the standard deviation of the differences between paired values.

η^2 and Cohen's d can be directly obtained in the free software JASP (Version 0.17.1) (JASP, 2023). Therefore, this software is used to analyze differences in physiological indicators.

The whole process of statistical analysis for PAD and physiological response data is shown in Figure 3.10.

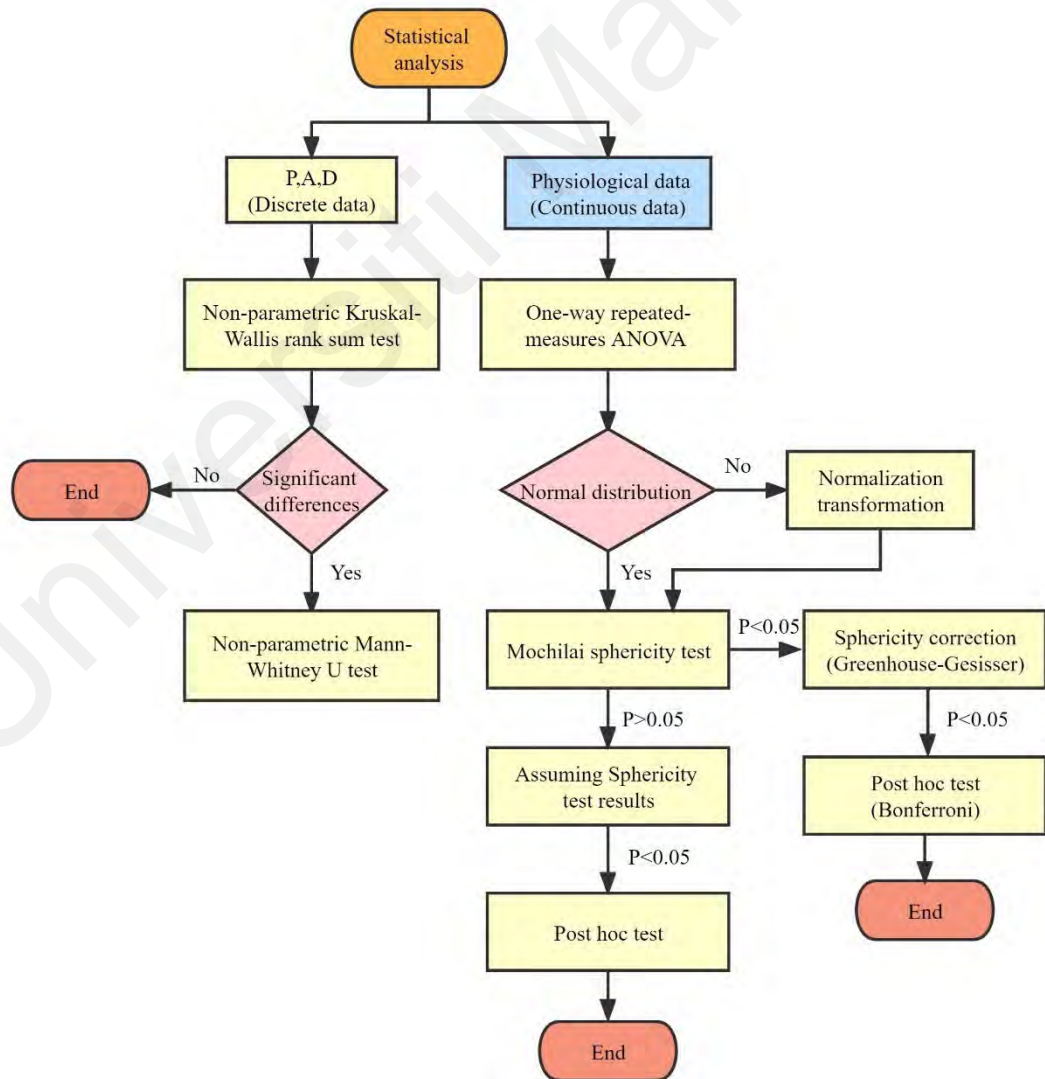


Figure 3.10: The statistical analysis process

3.9.1.3 Personal difference

One-way repeated Analysis of Variance was conducted to examine the effects of risk levels on emotional and physiological responses. ANOVA can determine whether the majority of the variation in a metric is due to between-group factors (risk level) or within-group factors (different subjects). If the former is significantly larger enough than the latter, it can be concluded that the differences are mainly caused by between-group factors (also known as controlled variables). However, this method cannot determine whether there are differences within groups. Therefore, a 2-step Hierarchical Linear Modeling (HLM) is employed to investigate individual differences. This method has demonstrated successful application in examining significant individual variations in Electrodermal Activity (EDA) while controlling for risk levels in the construction field (B. Choi et al., 2019). Besides, it was also utilized to ascertain the effects of trait dominance on risk-proneness, except for the gender factor (Demaree et al., 2009). HLM is suitable for analyzing data with a hierarchical structure (Suraji et al., 2001). In the data structure of this study, the dependent variables were emotional and physiological response indicators, and the influencing factors were risk level and individual differences. Everyone's data would be recorded at four risk levels so that the data points can be defined as Level 1 (each risk level) and Level 2 (each subject). Thus, a two-level HLM model can be used to determine whether there was individual difference. According to the calculation steps (Heck et al., 2013), the initial independent variables are included in the regression model, and the subsequent variables are added to the variables previously included in the previous step. Therefore, the essence of this method is to construct the regression model by gradually adding independent variables and observing the effect of each new variable on the dependent variable. Suppose this effect, i.e., the explained variance of the dependent variable, increases significantly after adding the new independent variable. In

that case, the new independent variable can significantly affect the dependent variable. In other words, the new variable significantly contributes to the variation-dependent variable.

3.9.2 Machine Learning

The results of statistical differences in physiological and emotional reactions reflecting perceived risk may not directly apply to determining an individual's perceived risk (B. G. Lee et al., 2021). So, it is necessary to mine deep relationships between reactions and perceived risk. The ability of ML is considered to recognize patterns hidden in data (Carleo et al., 2019). Moreover, unlike conventional statistical analysis models, machine learning demonstrates superior effectiveness when dealing with collected data under minimal assumptions and complicated nonlinear interactions (Ij, 2018). Several commonly used supervised machine learning models are trained and validated to achieve the research objectives using a broad range of features extracted from EDA, HRV, ST and PD to recognize cognitive risk. Figure 3.11 displays the process of machine learning.

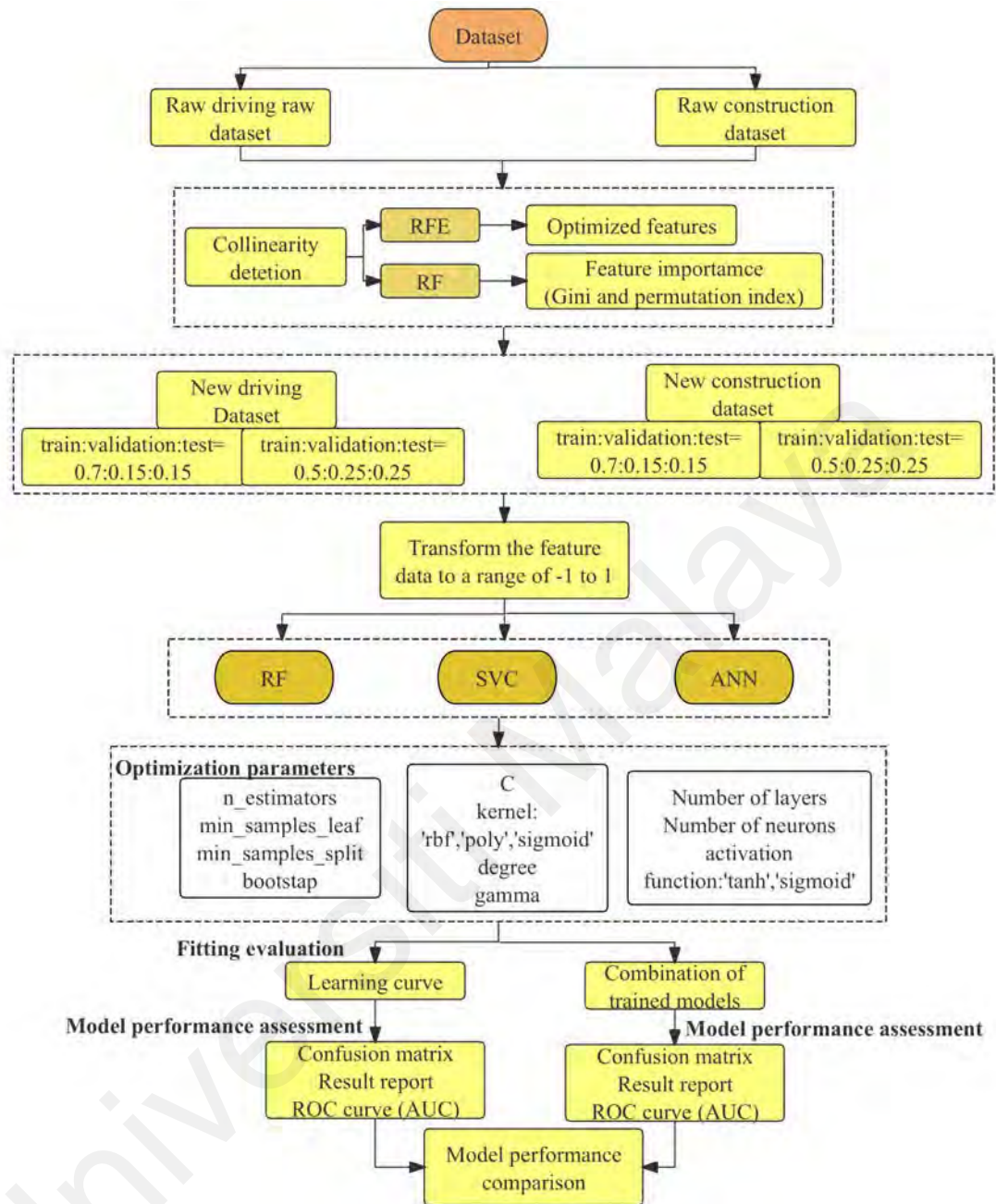


Figure 3.11: The machine learning process for classifying perceived risk levels and determining feature importance

3.9.2.1 Scaled raw dataset

In machine learning, the input data may have inconsistent scales, such as the range of emotion dimension being 1-9, EDR ranging from 0.1 to 1.2, and HR ranging from 55 to 115. Therefore, it is necessary to transform all data to a range of -1 to 1 by MinMaxScaler.

3.9.2.2 Collinearity detection and feature importance

20 indicators, including 17 physiological and 3 emotional indices, were input vectors. In order to assess the sensitivity of each indicator in the classification process, feature importance provided by RF can be employed. A higher value indicated greater importance, while a lower value suggests lesser importance. Feature importance reduced within the field of feature engineering in machine learning. It is calculated based on the influence of a particular indicator on the output results during the RF model computation process. It can be considered an ancillary output of the RF model construction process.

However, it is important to note that a prerequisite for calculating feature importance is that the variables involved must be mutually independent. Multicollinearity among variables can significantly affect the calculation result, potentially resulting in outcomes that deviate significantly from reality (Chan et al., 2022; Rawal & Ahmad, 2021). So, the initial step involves detecting collinearity among the 20 variables. It can be achieved by calculating the correlation coefficient of paired vectors to assess the correlation between variables. Therefore, before using RF to compute feature importance, it is essential to qualitatively assess the correlations among indicators through the correlation matrix and quantitatively diagnose multicollinearity.

Feature selection involves identifying and eliminating irrelevant and redundant features as much as possible (Yu & Liu, 2004). Furthermore, it helps mitigate the overfitting problem by simplifying and generalizing the model, thereby enhancing its accuracy (Sarker et al., 2020). Recent developments in machine learning and optimization have shown better results than conventional statistical methods (Chan et al., 2022). Recursive feature elimination (RFE), a popular technique based on machine learning, is used for selecting independent features. RFE iteratively removes the weakest feature during model fitting until the specified number of features is reached. This gradual

elimination reduce dependencies and collinearity (Pedregosa et al., 2011). Subsequently, the importance of the selected feature is calculated.

After selecting features, Mean Decrease Impurity (MDI) and Permutation Importance (PI) were used to represent feature importance in this study. They can be calculated with RF during classification. Specially, Gini importance is a common measure of MDI. It quantified the reduction in the Gini index when each feature is used to split the samples in a decision tree. The Gini index reflects the sample impurity level, so the Gini importance assesses how each feature contributes to impurity reduction. Permutation Importance (PI) measures feature importance differently. It involves randomly shuffling each feather and assessing the resulting change in model prediction accuracy. If a feature significantly affects performance, it is considered necessary; otherwise, it is deemed less critical (scikit-learn, 2023). MDI and PI yield larger values for more important features, indicating their higher significance.

3.9.2.3 Split Dataset

The two datasets are randomly divided into three subsets: one for model training, one for model validation, and the last one for model testing. Typically, the training set comprises 70% or 80% of the data, while the testing set consists of 30% or 20%. However, physiological and emotional reacting data have unique characteristics where most indicators do not follow a non-normal distribution. It poses significant challenges in parameter optimization, as accurate prediction results heavily rely on the precondition that training and test data share the same distribution.

The two datasets were separated using two approaches to address this issue. The first approach involved allocating 70% of the data for training, 15% for validation, and 15% for testing. The second approach allocated 50% for training, 25% for validation, and 25% for testing. By evaluating the results of the same model on these two datasets, it becomes

possible to assess the impact of inconsistent distributions and further analyze their influences.

3.9.2.4 Model selection

Figure 2.5 presents an overview of the three most widely employed machine learning algorithms, namely ANN, SVM, and RT, for analyzing physiological signals within the human factor domain. Additionally, in a study using machine learning to classify perceived risk (B. G. Lee et al., 2021), SVM with the Gaussian kernel function demonstrated superior performance compared to the K-Nearest Neighbors (KNN) and DT models. The study utilized a combination of physiological signals to classify perceived risk. So, the three algorithms were selected in this research. It is important to note that the input vectors consist of 18 physiological and three emotion dimensions, while the output results correspond to four risk levels (1 to 4). This setup clearly indicates a typical classification problem, which necessitates the use of these three algorithms for classification purposes. Support Vector Classification (SVC) was then employed for classification, replacing the traditional SVM. The kernel function of SVC included ‘poly’ and ‘sigmoid’ besides ‘rbf’ (representing Gaussian function) in the present study. Additionally, Random Forest (RF) can be seen as an ensemble learning algorithm of decision trees, and in this case, it is used as a substitute for Decision Tree (DT).

3.9.2.5 Classification

Parameter optimization: RF, SVC, and ANN were selected to utilize the training dataset for model training. This process optimizes the parameters of each model, ensuring improved accuracy on the training dataset. For machine learning algorithms, hyperparameters cannot be obtained through model training. Standard methods for determining hyperparameters include grid search, random search, and automated parameter optimization. Grid search and random search can often get trapped in local optima due to their search strategies. Therefore, this study used Hyperopt, a Python

library, for automatic hyperparameter optimization. It is designed for hyperparameter optimization. It uses Bayesian optimization to adjust parameters and allows obtaining the best parameters for a given model. It can optimize models with hundreds of parameters within a relatively large range. So, it is introduced to conduct parameter optimization.

The performance of the three models was also assessed using the validation dataset. Figure 3.11 indicates the parameters to be optimized for each model.

Evaluation fitting performance: The fitting performance of each model needs to be evaluated initially, ensuring that neither overfitting nor underfitting occurs. Overfitting indicates that the trained model performs perfectly on the training dataset but poorly on the validation data. On the other hand, underfitting suggests that the trained model must be able to accurately predict the output or capture the underlying relationship between input vectors and risk levels. Both scenarios are undesirable and require evaluation before assessing the models. Excellent fitting performance is characterized by increasing accuracy in both the training and validation datasets as the training progresses, with the gap between them gradually decreasing or remaining stable.

Model performance assessment: This research aims to address a classification problem and evaluate or compare the performance of different models using two approaches to find the best model.

Firstly, the models were evaluated on a test set using a confusion matrix, precision, accuracy, recall, and F1 score. These metrics are derived from the confusion matrix, but the visual representation of the confusion matrix is more effective in providing a comprehensive understanding of classification performance. Furthermore, these metrics provide insights into the performance of the models from different perspectives, with higher values indicating better performance.

Secondly, Receiver Operating Characteristic (ROC) curves were utilized. Since the problem is multi-class, ROC curves are generated by comparing each class against the

rest, and the corresponding area under the curve (AUC) values are calculated. Generally, a ROC curve closer to the upper-left corner of the coordinate system and a higher AUC value indicate better model performance.

3.9.2.6 Integration of trained models

It is important to emphasize that the first and last situations observed in driving and confirmed experiments display significant differences. The former is relatively safe, while the latter represents the highest level of risk. However, the two middle situations share similar risk levels, posing a significant challenge for machine learning models in accurately classifying them. Nevertheless, the size of available datasets limits the utilization of deep learning. Therefore, the model integration is approached from two different perspectives.

In the first approach, the output is obtained by inputting a 15% proportion of the validation data to the three pre-trained models. The term P_{ij} is used to represent the output probability, where ‘ i ’ corresponds to the model (ranging from 1 to 3), and ‘ j ’ denotes the category index (ranging from 1 to 4). Essentially, P_{ij} signifies the probability of the i -th model classifying a sample into the j -th category. When P_{ij} exceeds 0.5 for a particular category, the sample is classified accordingly. This process, Following Formula 3-1, this process accumulates to yield n_{ij} , representing the count of samples assigned to the j -th category by the i -th model. In total, for the three models and four classes, a matrix N_{ij} is obtained with dimensions 3×4 .

$$N_{ij} = [n_{ij}] \left\{ \begin{array}{l} n_{ij} = n_{i-1j-1} + 1, p_{ij} > 0.5 \\ i = 1,2,3; j = 1,2,3,4 \end{array} \right. \quad (3.6)$$

Next, the weight coefficients of the i -th model in the j -th category are calculated according to Formula 3-7.

$$w_{ij} = n_{ij} / \sum_{i=1}^3 n_{ij} \quad (3.7)$$

Finally, 15% of the test data is input into the three pre-trained models, applying the same processing method as the validation set, yielding corresponding results denoted as M_{ij} . According to Formula 3-8 the probability values of the test set for the j -th class are calculated as P_j . The classification result, as described in Formula 3.9, is determined by selecting the 'j' value corresponding to the maximum probability.

$$P_j = \sum_{i=1}^3 m_{ij} \cdot w_{ij} \quad (3.8)$$

$$\max_{j=1}^4 (P_j / \sum_{j=1}^4 P_j) \quad (3.9)$$

In such cases, combining three trained models becomes valuable, allowing us to maximize the strengths of each model across different classes. This may significantly improve the recognition ability of each category, especially for risk situations with minimal variations. A graphical representation is presented in Figure 3.12 to clarify the computational procedure described above.

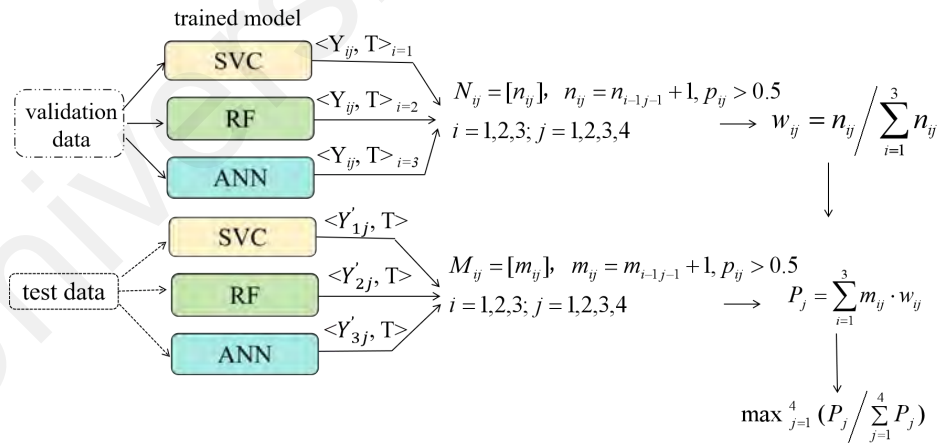


Figure 3.12: The integration of the computational process of trained models through weights calculation for different categories

In the second approach, the validation dataset is fed into the three pre-trained algorithms to obtain probability values for each class, following the same weight adjustment. Subsequently, these output results serve as new input data for training an additional Random Forest (RF) model. The reasons why choosing random forest in the

stacking technique was its robustness and stability, diverse decision boundaries and reduced overfitting (Breiman, 2001; Zhou, 2012). In essence, a new model is stacked on top of the existing three, employing the concept of stacking. Finally, the test data is input into the newly trained RF model and compared its predictions with the other algorithms. This computational process is visualized in Figure 3.13.

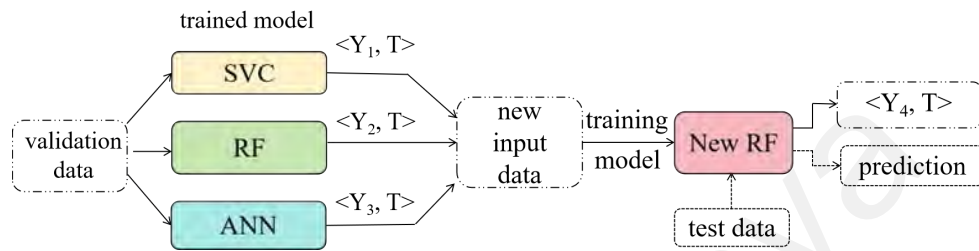


Figure 3.13: The integration of the computational process of trained models through the stacking technique

Both approaches are based on the three pre-trained base models, aiming to maximize the classification advantages of different models across various classes. The objective is to enhance the performance of model combinations, particularly in cases where classification is more challenging. However, these approaches differ in their methods: one involves increasing the weights of models that correctly classify specific categories, while the other leverages the stacking concept to enhance the overall performance of the classification models.

3.10 Summary

This chapter explains the whole experimental design, detailed implementation process, and data analysis methods.

Considering the first objective, four driving scenarios were selected as experimental stimuli, measuring physiological and emotional responses associated with risk perception through EDA, ECG, pupil diameter, skin temperature, and the PAD emotional model. Subsequently, statistical analysis methods were employed to examine physiological and emotional responses. The objective was to discern whether these differences stemmed from individual differences or contextual risks. Notably, this study deliberately included

two risk scenarios with closely matched risk levels, allowing for a more nuanced analysis compared to previous research.

For the second objective, we applied machine learning algorithms, building upon the foundation of statistical analysis, to construct models for uncovering patterns of variation in physiological and emotional responses to risk levels. We utilized three individual learning algorithms (RF, SVC, and ANN), along with the combinations of these three algorithms, to identify the optimal machine learning models.

Finally, an experiment in a confirmed construction setting was designed to validate the research outcomes for both objectives, employing a similar experimental procedure and data analysis methods. It allowed for a comparison of the results of statistical analysis and machine learning classification between the initial and confirmed experiments.

Universiti Malaysia

CHAPTER 4: STATISTICAL ANALYSIS OF PHYSIOLOGICAL AND EMOTIONAL RESPONSES

4.1 Introduction

The experimental design elicited the subjects' physiological and emotional responses based on driving stimuli. It was aimed to explore the neural regulatory mechanisms of the human body in risky situations. Statistical methods were conducted to analyze these responses comprehensively and identify indicators with significant differences.

The analysis had two primary objectives: firstly, to determine if there were statistical variances in specific emotional and physiological parameters, and secondly, to determine instances of significant differences in measurement indicators and whether these variances were linked to varying risk levels or individual dissimilarities.

4.2 Statistical Analysis of PAD

In the initial experiment, 60 subjects participated, but the physiological data of 6 participants still needed to be excluded. Consequently, an analysis was performed using data from 54 samples in driving situations. Table 4.1 displays the medians, 25th and 75th quartiles of P, A, and D, along with the key findings from the Kruskal-Wallis test.

Table 4.1: Non-parameter test and pair comparison results of P, A, and D in driving

Variables	P				A				D			
Group	1	2	3	4	1	2	3	4	1	2	3	4
Media	5	7	7	9	5	4	3	1	7	5	5	1
25%	5	6	7	9	4	3	2	1	5	4	3	1
75%	6	7	8	9	5	4	4	1	7	6	5	2
H	79.674**				82.984**				97.397**			
Cohen's f	0.073				0.077				0.094			
Pair comparison (Mann-Whitney U test)	test	Cohen's d			test	Cohen's d			test	Cohen's d		
	1-4, t=363.5**	1.47			1-4, t=2655* *	1.89			1-4, t=2795* *	2.68		
	2-4, t=329**	1.24			2-4, t=2591* *	1.47			2-4, t=2682* *	2.11		

Table 4.1, continued

	test	Cohen's d	test	Cohen's d	test	Cohen's d
	3-4, t=473**	0.88	3-4, t=2475* *	1.12	3-4, t=2526* *	1.59
	1-3, t=816.5**	0.82	1-3, t=2024.5 **	0.62	1-3, t=2083.5 **	0.85
	2-3, t=1105.5*	0.46	1-2, t=1907*	0.50	1-2, t=1951*	0.61
	1-2, t=1087.5*	0.41	2-3 t=11702. 5P=0.24 9		2-3 t=1669 P=0.376	

Note: *, $p < 0.05$, **, $p < 0.01$

Figure 4.1 plots the values of three dimensions to depict the changing pattern of P, A, and D about risk levels. The plot revealed a consistent increase in P as the risk level escalated, while A and D demonstrated a declining trend. Notably, when the risk levels were relatively close (2 and 3), the median values of P and D were equivalent, with A differing only by 1. Significantly, in the highest-risk scenario (4), the values peaked.

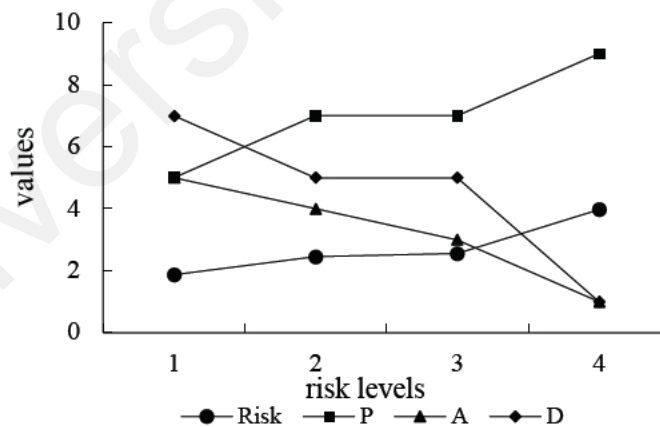


Figure 4.1: The changing trend of P, A, and D with risk ranks in driving

Through the Kruskal-Wallis test, all three dimensions (P, A, and D) exhibited significant differences among the four situations.

It can be observed that the effect size of D (0.094) was more extensive than that of P (0.073) and A (0.077). However, all three effect sizes were below 0.1, indicating a small effect size. Additionally, pairwise comparison analysis (Mann-Whitney U Test) revealed

significant differences in all three dimensions between the relatively safe context (1) and the risky context (4), between the near-loss contexts (2 and 3) and the risky context (4), and between the safe context (1) and the near-loss context (3). Moreover, all of these effect sizes exceeded 0.8, indicating large effect sizes, except for the effect size of A (0.62) between contexts 1 and 3, which falls within the medium effect size range. When the difference in risk degrees was slight, such as between contexts 2 and 3, there were no significant differences for A and D. At the same time, P exhibited a small effect size difference (0.41) between these two scenarios.

4.3 Statistical Analysis of Physiological Data

4.3.1 Statistical analysis of EDA

EDR, as the momentary skin conductance response, showed a continuously increasing trend with the rise of risk level, ranging from 0.16 to 0.34. In construct, EDL, as the baseline skin conductance, exhibited a pattern of first decreasing and then increasing at the highest level of risk. Both indicators did not follow a normal distribution and violated the assumption of sphericity ($\chi^2 = 28.605$, $P < 0.001$ for EDL, and $\chi^2 = 17.074$, $P = 0.004$ for EDR). Because of the result of the correction coefficient (Epsilon = 0.575 < 0.75 for EDL, Epsilon = 0.613 < 0.75 for EDR), the Greenhouse-Geisser corrections were applied to adjust the degrees of freedom. According to the test statistics (EDL: $F(2.411, 127.786) = 3.991$, $p = 0.015$, $\eta^2 = 0.070$), EDL showed significant differences in risk levels. EDR were also significantly different (EDR: $F(2.481, 131.471) = 17.143$, $p < 0.001$, $\eta^2 = 0.244$) through the same analyzing process with EDL.

Furthermore, the results of post-hoc analysis of Bonferroni showed EDL was significantly larger in high-risk level (4) than low-risk (3) with $P < 0.05$, whereas no significant differences manifested among other risk levels. The post hoc test with a Bonferroni correction applied for EDR demonstrated that there were significant differences among the four pairs of risk levels (1-3, 1-4, 2-3, 2-4), and the effect size was

largest between risk levels 1 and 4 (Cohens'd=-0.636). Conversely, the degree of difference between risk levels 2 and 3 (Cohens'd = -0.384) was the smallest. The magnitude of differences between the other two pairs fell between the mentioned above, with Cohens'd values of -0.499 (1-3) and -0.521 (2-4), respectively.

Table 4.2: Description and repeated-measure ANOVA analysis results of EDA in driving

Variables	EDL/ μs				EDR/ μs			
Group	1	2	3	4	1	2	3	4
Mean	0.26	0.23	0.18	0.33	0.16	0.19	0.30	0.34
Std.	0.69	0.60	0.55	0.65	0.17	0.31	0.34	0.28
Test of Sphericity	$\chi^2=28.605$, $P<0.001$				$\chi^2=17.074$, $P=0.004$			
Within Subjects Effects	F (2.411,127.786) =3.991, $p=0.015$				F (2.481,131.471) =17.143, $p<0.001$			
η^2	0.070				0.244			
Post Hoc Comparisons	T test		Cohens'd		T test		Cohens'd	
					1-3, -0.14*	-0.499		
	3-4, -0.15*		-0.247		1-4, -0.18**	-0.636		
					2-3, 0.11*	-0.384		
					2-4, -0.15*	-0.521		

Notes: *, $p<0.05$, **, $p<0.01$

4.3.2 Statistical analysis of HRV

The original ECG data underwent preprocessing using the methods and parameters outlined in Table 3.3. This preprocessing involved filtering, noise removal, and R-peak extraction. Subsequently, ErogLAB software was utilized to compute time-domain, frequency-domain, and nonlinear indices from the processed data.

The time-domain indicators of HRV mainly included HR, SDNN, RMSSD, and SDD. The mean and standard deviation of the four indicators are listed in Table 4.3. The indicators did not consistently increase or decrease as the risk level increased. This inconsistency may be attributed to the changes in their values observed in the third scenario, which deviated from the original trend. An interesting observation was that in the context of the highest risk, the values of these four time-domain indices were nearly the lowest besides SDNN.

The within-subjects test did not reveal significant differences for SDNN, RMSSD and SDSD. However, significant differences were shown for HR. HR data followed a normal distribution and passed the sphericity test. A repeated measures analysis of variance confirmed the presence of a significant difference. Post hoc tests revealed a significant difference between 1 and 4, indicating that HR under the highest risk level was significantly lower than the relatively safe condition. HR in near-loss situation 2 was 3.89 less than the HR value in the safe scenario 1, which was statistically significant.

Table 4.3: Description and repeated-measure ANOVA analysis results of HRV time-domain indicators in driving

Variables	HR/bpm				SDNN/ms			
Group	1	2	3	4	1	2	3	4
Mean	79.57	75.69	76.00	74.83	98.12	61.65	69.72	68.35
Std.	14.99	14.51	16.63	14.38	258.80	100.27	131.45	105.18
Test of Sphericity	$\chi^2 = 7.334, P = 0.197$				$\chi^2 = 151.618, P < 0.001$			
Within Subjects Effects	F (3,159) = 6.038, p < 0.001				F (1,336,70.796) = 0.835, p = 0.396			
η^2	0.102							
Post Hoc Comparisons	T test		Cohens'd		T test		Cohens'd	
	1-2, 3.89*		0.273					
	1-4, 0.85**		0.333					
Variables	RMSSD/ms				SDSD/ms			
Group	1	2	3	4	1	2	3	4
Mean	117.61	73.75	86.32	67.62	119.12	76.47	90.89	68.81
Std.	379.19	170.79	196.08	112.66	385.25	179.67	208.56	113.61
Test of Sphericity	$\chi^2 = 138.450, P < 0.001$				$\chi^2 = 128.100, P < 0.001$			
Within Subjects Effects	F (1,325,70.208) = 0.736, p = 0.430				F (1,359,72.009) = 0.692, p = 0.451			

Note: *, $p < 0.05$, **, $p < 0.01$

There are five indices in the frequency-domain analysis of HRV. Five frequency-domain indices were obtained after preprocessing the original data and performing a fast Fourier transform (FFT) on the time-domain signal (ULF, VLF, LF, HF and LF/HF). The description and statistical results of the five indicators are listed in Table 4.4.

Table 4.4: Description and repeated-measure ANOVA analysis results of HRV frequency-domain indicators in driving

Variables	ULF				VLF			
Group	1	2	3	4	1	2	3	4
Mean	0.87	0.46	0.38	0.51	17.60	7.74	5.96	8.48
Std.	0.94	0.48	0.37	0.49	14.03	6.59	4.81	6.82
Test of Sphericity	$\chi^2=36.463$, $P<0.001$				$\chi^2=41.814$, $P<0.001$			
Within Subjects Effects	F (1,994,105.670) =7.026, $p=0.001$				F (1,909,101.189) =20.720, $p<0.001$			
η^2	0.117				0.281			
Post Hoc Comparisons	T test		Cohens'd		T test		Cohens'd	
	1-2, 0.41*		0.676		1-2, 9.90**		1.124	
	1-3, 0.49**		0.806		1-3, 11.68**		1.326	
	1-4, 0.37*		0.602		1-4, 9.16**		1.041	
Variables	LF				HF			
Group	1	2	3	4	1	2	3	4
Mean	39.71	38.93	35.42	40.50	41.77	52.87	58.23	50.52
Std.	21.10	24.10	20.91	20.79	25.46	29.06	23.96	25.41
Test of Sphericity	$\chi^2=1.017$, $P=0.961$				$\chi^2=1.215$, $P<0.001$			
Within Subjects Effects	F (3,159)=0.895, $p=0.445$				F (3,159)=5.745, $p<0.001$			
η^2					0.098			
Post Hoc Comparisons	T test		Cohens'd		T-test		Cohens'd	
					1-2, -11.09*		-0.426	
					1-3,-16.46**		-0.632	
Variables	LF/HF							
Group	1	2	3	4				
Mean	2.17	1.71	0.98	1.46				
Std.	3.39	2.38	1.07	1.63				
Test of Sphericity	$\chi^2=32.784$, $P<0.001$							
Within Subjects Effects	F(2.130,112.874)=2.73 9, $p=0.062$							

Notes: *, $p<0.05$, **, $p<0.01$

To facilitate the visualization of the trends in the data changes, the mean values of these indicators concerning changes in risk level were plotted in Figure 4.2.

Firstly, HF and LF had the highest proportion, accounting for approximately 40% to 60%. Secondly, all other frequency-domain indicators, except for HF, showed a gradual

decline, followed by an increase in the last scenario. HF showed an opposite pattern, with an initial increase followed by a decrease. Thirdly, the other three frequency-domain indices (ULF, VLF and HF), apart from LF and LF/HF, showed significant differences at different levels of risk, and the VLF reached the most significant degree of difference ($\eta^2=0.281$).

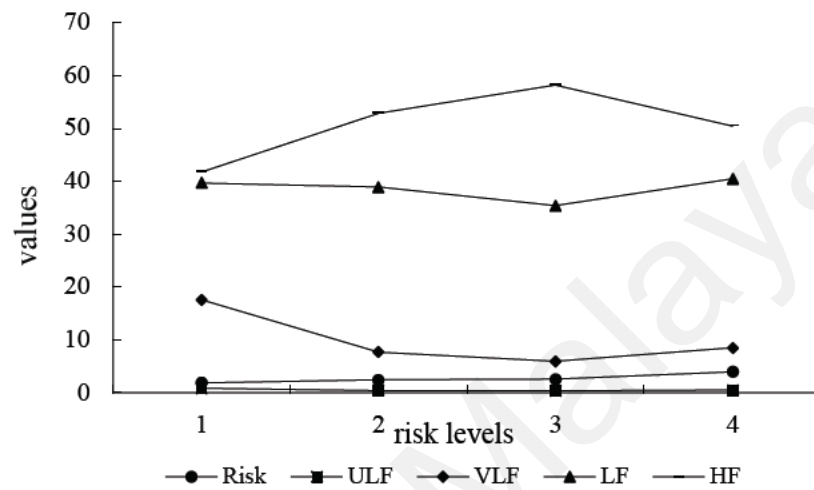


Figure 4.2: Mean values of HRV frequency-domain indicators with risk level

In HRV analysis, the Poincaré plot can help analysts evaluate the complexity and nonlinear characteristics of heart rate variability, providing more detailed information to understand better the characteristics and physiological mechanisms of heart rate variability. Four commonly used indicators include SD1, SD2, A++ and B--. The statistical results for the four indicators are listed in Table 4.5.

Table 4.5: Description and repeated-measure ANOVA analysis results of HRV nonlinear indicators in driving

Variables	SD1				SD2			
	1	2	3	4	1	2	3	4
Group	84.23	54.07	64.27	48.66	106.75	64.58	70.48	82.19
Mean	84.23	54.07	64.27	48.66	106.75	64.58	70.48	82.19
Std.	272.41	127.05	147.47	80.33	246.01	68.17	115.96	127.80
Test of Sphericity	$\chi^2 = 128.100, P < 0.001$				$\chi^2 = 141.918, P < 0.001$			
Within Subjects Effects	F (1.359, 72.009) = 0.692, p = 0.451				F (1.282, 66.676) = 0.802, p = 0.402			

Table 4.5, continued

Variables	A++				B--			
Group	1	2	3	4	1	2	3	4
Mean	2.87	3.93	2.69	4.19	2.44	3.06	2.70	3.37
Std.	1.32	2.27	1.48	2.04	1.21	1.40	1.77	2.01
Test of Sphericity	$\chi^2 = 21.389, P < 0.001$				$\chi^2 = 7.453, P < 0.189$			
Within Subjects Effects	F (2,429,128.762) = 11.740, p < 0.001				F (3,159) = 5.651, p = 0.001			
η^2	0.181				0.096			
Post Hoc Comparisons	T test		Cohens'd		T test		Cohens'd	
	1-2, -1.06*		-0.606		1-4, -0.93**		-0.615	
	1-4, -1.32**		-0.744		3-4, -0.67*		-0.437	
	2-3, 1.24**		0.732					
	3-4, -1.50**		-0.861					

Notes: *, $p < 0.05$, **, $p < 0.01$

SD1 and SD2 were statistical indicators that reflect heart rate variability. SD1 was usually used for representing the short-term variability of heart rate during the respiratory cycle, mainly reflecting the variability in the high-frequency range (0.15-0.4 Hz). The larger the value, the more sympathetic and parasympathetic nervous system regulation is balanced. SD2 represented the overall heart rate variability over a long period, mainly reflecting the variability in the low-frequency range (0.04-0.15 Hz) (Kemp et al., 2017). The larger the value, the more regulation between SNS and PNS is coordinated.

The nonlinear physiological analysis of HRV can be summarized as follows. In usual situations, the values of SD1 and SD2 were relatively higher (84.23 and 106.75, respectively) than in any other dangerous situations. However, in situations where the danger had already occurred but did not result in negative consequences (situations 2 and 3), both SD1 and SD2 values were lower than in safe scenes. When the risk reached its highest level, meaning the presence of danger and the consequence of injury to individuals, SD1 decreased (from 64.27 to 48.66), while SD2 increased (from 70.48 to 82.19). However, both SD1 and SD2 indicators did not show statistically significant differences.

A⁺⁺ represents the strength or complexity of certain nonlinear features, while B⁻⁻ represents the heterogeneity or regularity of these features in heart rate variability. A high A⁺⁺ value indicates intense or complex nonlinear features, while a low B⁻⁻ value suggests more regular or less heterogeneous nonlinear features. The A⁺⁺ and B⁻⁻ trends were similar, but A⁺⁺ consistently had higher values than B⁻⁻ in all situations. Particularly in the last and most risky situation, both A⁺⁺ and B⁻⁻ reached their maximum value.

Moreover, A⁺⁺ and B⁻⁻ showed significant differences through repeated-measure ANOVA. However, the effect size of A⁺⁺ ($\eta^2=0.181$) was greater than B⁻⁻ ($\eta^2=0.096$). Regardless of whether it was based on the number of pairwise groups with significant differences or Cohens' d values that represent the magnitude of effect sizes, A⁺⁺ was consistently higher than B⁻⁻.

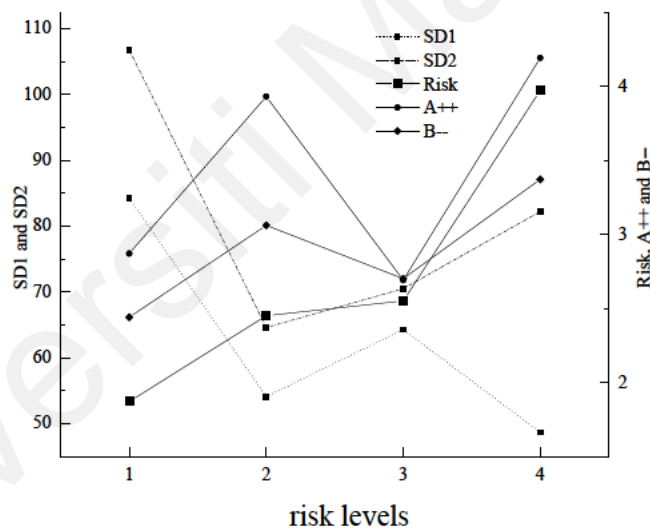


Figure 4.3: Mean values of HRV nonlinear indicators with risk level

4.3.3 Statistical of PD and ST

Pupil diameter is often utilized to assess the mental workload (Marquart et al., 2015). Skin temperature is considered an indicator of SNS, and its growth may indicate the strength of the activities of SNS (Kim et al., 2004). However, upon reviewing Table 4.6, it can be observed that both presented a minor fluctuation among the four situations and had no statistical differences.

Table 4.6: Description and repeated-measure ANOVA analysis results of PD and ST in driving

Variables	PD				ST			
	1	2	3	4	1	2	3	4
Group	1	2	3	4	1	2	3	4
Mean	4.30	4.10	4.12	4.18	33.78	34.23	33.89	34.39
Std.	0.69	0.60	0.65	0.69	1.72	1.47	1.72	2.73
Test of Sphericity	$\chi^2 = 23.172, P < 0.001$				$\chi^2 = 27.348, P < 0.001$			
Within Subjects Effects	F (2.332, 123.614) = 0.735, p = 0.061				F (2.193, 116.225) = 1.073, p = 0.350			

4.4 Discussion of Emotional and Physiological Responses

The study of subjective risk perception from the perspective of individuals' intrinsic physiological responses has emerged as a new trend. Furthermore, when considering the interrelation and mutual influences between psychological and physiological systems, it becomes evident that these changes can provide valuable information for understanding the process of perceived risk.

4.4.1 Discussion of Emotional Responses

P, A and D are necessary and sufficient for the description of any emotional state. Since it enables the representation of emotional states of an intelligent entity in a three-dimensional space (Rincon et al., 2017), it has been used successfully in several relevant emotion expression areas, such as advertising effect assessment (Morris & McMullen, 1994), consumer choice (Platania et al., 2016), web design (Verkijika & De Wet, 2019) and emotion recognition (Häring et al., 2011; Hartmann et al., 2013).

4.4.1.1 Emotional state changes

As the level of risk increased from relative safety to higher risk, individuals underwent an emotional transition characterized by a shift from a neutral and calm state to a state of unhappiness, from a sense of normalcy to heightened arousal, and from a feeling of control to a perception of losing control. When the risk level peaked, the three experiences reached their maximum intensity. It was through the combined contribution of the three

dimensions that emotional responses under dangerous scenarios can transform into valuable information (Rincon et al., 2017).

Notably, the dominance dimension showed a negative relationship with risk perception. This finding was consistent with survey results conducted in a chemical factory and nuclear power plant (Mbaye & Kouabenan, 2013). Additionally, the observations were aligned with the assumption proposed by Megias et al. (2018). In other words, risk perception was not limited to low and high situations but exhibited multi-level distinctions based on the magnitude of risk.

4.4.1.2 Differentiation ability among three dimensions

Although the three emotional dimensions showed significant differences across four levels of risk in driving scenarios, their ability to differentiate risk varies was different. In the driving context, D exhibited the highest value regarding Cohens'd, followed by A and P. So, the performance of D should not be underestimated. The possible reason maybe that the subjective judgement of risk relied on the controlling degrees of their involved environment (Flin & Mearns, 1994). In a study on the influencing factors of adventurous behavior, D demonstrated a remarkable capacity to accurately predict individuals' inclination toward taking risks (Demaree et al., 2009). As one dimension of emotional space, D was widely recognized as an essential component that aided in the categorization of emotions (Ozel et al., 2019).

4.4.2 Discussion of Physiological Responses

4.4.2.1 The changes of EDA

EDR increased when individuals became aware of the presence of danger. This finding was consistent with previous research. For example, studies have shown that EDR exhibits higher values in hazardous situations (such as crashes or near misses) than in safe driving conditions (Tagliabue et al., 2019). In other driving situations, it increased in the presence of danger compared to normal driving conditions (Barnard & Chapman, 2016;

Perello-March et al., 2022). Similarly, pilots have been found to have higher EDR values during incident phases compared to regular flight phases (Lutnyk et al., 2023). Furthermore, EDR usually increases as the perceived risk become greater or the risk level rises. For example, in driving scenarios, higher-risk situations such as navigating roundabouts have been associated with more significant EDR responses, likely due to the activation of heightened arousal levels, which can also be observed through decreased driving speed (Distefano et al., 2022).

Although the effect size of the EDL component was lower compared to EDR, it can still reflect the baseline level of arousal in different situations. In a measuring EDA experiment conducted on construction sites, EDL showed no significant difference between high and low-risk activities (B. Choi et al., 2019), which seemed different from the present research. A meticulous analysis revealed that the disparities in EDL observed in this study stemmed from high-risk situations 3 and 4 rather than low-risk situation 1 and the other three hazardous scenarios. Viewed in this light, the conclusions drawn from both studies were, in fact, congruent. It was evident that the variations in EDL differed from EDR, as individuals exhibited significant differences in their physiological arousal levels only when exposed to distinct risk situations, thereby reflecting divergent physiological preparedness in response to risk situations.

EDR is a more sensitive marker to distinguish different risky situations than EDL. Given that the effect size of EDR (0.244) exceeded that of EDL (0.070), EDR not only functioned as a physiological marker for discriminating between low-risk and high-risk situations but also demonstrated a more robust and consistent capability to differentiate various levels of risk, even in situations encompassing both threat presentation and adverse outcomes. Because statistically significant differences in EDR were observed in driving scenarios that involved near-miss incidents (scenarios 2 and 3) due to different causes, these findings not only confirmed the well-established ability of EDR to

differentiate between high-risk and low-risk or safe situations but also suggested its potential for distinguishing between scenarios with less variation in risk levels. Therefore, EDR, as a robust marker of bodily arousal and expression of emotion, holds significant potential for mapping cognitive perceptions of risk levels (Horn et al., 2020).

The EDR sensitivity can be explained by its mechanism. When an external stimulus occurs, as long as it elicits emotional arousal in the subjects, the brain sends signals to activate the eccrine sweat glands through the sympathetic branch of the autonomic nervous system (Dawson et al., 2007; Larkin, 2006). These glands secrete sweat, leading to an increase in skin conductance. Consequently, the skin conductance value goes up. This entire process happens within a few seconds: the rise in skin conductance starts 1 to 4 seconds after exposure to the stimulus and lasts for 1 to 3 seconds (Dawson et al., 2011; Measures et al., 2012). Therefore, the crucial factor here is that the sympathetic nervous system controls skin conductance and responds rapidly, which explains why it is a sensitive indicator.

4.4.2.2 The changes of *ECG*

The greater complexity observed in ECG variations compared to EDA may arise from their simultaneous regulation by both branches of the autonomic nervous system, namely the sympathetic (SNS) and parasympathetic (PNS) systems. The main findings of HRV can be summarized as follows:

(a) Analysis of time-domain indicators

The values of HR (79.57), SDNN (98.12), RMSSD (117.61) and SDDSD (119.12) were highest in the safer situation (1), indicating a prevailing dominance of the parasympathetic nervous system (PNS), which is characterized by a relatively calm and relaxed state. It was consistent with the common understanding that both components of the ANS influence heart rate, but the PNS had a faster and more predominant effect on resting HRV (Berntson et al., 1997).

When hazards began to occur, even at a low level, HR, influenced by both the PNS and SNS, showed a subtle change during the short-term measurement period within 3 minutes. However, the SDNN (61.65-69.72 for driving) and RMSSD (73.75-86.62 for driving) exhibited more considerable variations but were still lower than the baseline. The primary source of the variation of SDNN was parasympathetically-mediated RSA (Shaffer et al., 2014) and RMSSD was usually considered as a primary measure to estimate the vagally mediated changes (Penttilä et al., 2001; Saboul et al., 2013). These changes indicated that the balance of SNS and PNS in safer situations was disrupted. Additionally, there was a degradation of PNS and an increased activation of SNS happened for two moderately risky situations.

During high-risk situations, the values of HR (74.83), SDNN (68.35), RMSSD (67.62) and SDSD (68.81) were the lowest. Based on the balance disruption in low-risk situations, the degree of ANS imbalance continued to intensify. For instance, RMSSD, a primary time-domain measure in HRV, reflected beat-to-beat variations in heart rate. It indicated changes modulated by the vagus nerve (Shaffer et al., 2014). According to the polyvagal theory (Porges, 1995, 1997, 2001, 2003a, 2003b, 2007, 2009), the vagus nerve was divided into myelinated and unmyelinated branches. The older unmyelinated branch collaborates with the sympathetic nervous system (SNS) in situations of danger or life threat. It enables the organism to respond effectively to challenges (Kemp & Quintana, 2013).

(b) Analysis of frequency-domain indicators

In terms of measuring heart rate variability (HRV), there are two main components to consider: HF and LF. HF was seen as a sensitive indicator of PNS (Holzman & Bridgett, 2017). In the initial clips, HF had the lowest value, while RMSSD, another PNS-sensitive indicator, had the highest value in safety situations. It might appear contradictory, but it could be because HF was influenced by breathing patterns (Shaffer et al., 2014). It aligned

with the finding that the PNS was less active, as indicated by time-domain measures. Even though there was no significant difference in HF across all four situations, it tended to be lower in the highest-risk situations compared to near-loss situations. It could be because HF decreased during cognitive regulation of negative emotional stimuli (Kemp et al., 2017). It reflected the inhibitory effect of negative emotions on cognition, consistent with the reduced activity of the vagus nerve function. Compared to HF, LF has a more complex mechanism. Its changes follow a complex pattern. LF is typically thought to be influenced by the vagus nerve, the sympathetic nervous system, and pressure reflex mechanisms all at once (Berntson & Cacioppo, 1999). The interaction between the sympathetic and parasympathetic nervous systems seemed intricate in generating LF power (Shaffer et al., 2014). This complexity could explain the erratic fluctuations in LF indicators.

ULF and VLF are not the main components of the frequency domain and are not commonly used indicators for short-term measurements. They were generally utilized in measurement conditions spanning 24 hours (Kleiger et al., 2005). So, the diurnal rhythm oscillation of heart rate is the main source of ULF power. However, it's worth noting that in some studies, during short-term recordings with a stressor present, the activation of the SNS can push the heart rate into the lower part of the low-frequency band, resulting in the emergence of very-low-frequency (VLF) components (Shaffer et al., 2014). It aligned with the findings in this study. VLF was observed to be higher in the most dangerous situation (4) than in the moderate danger situations (2 and 3). It further confirmed the activation of the SNS and the occurrence of emergency responses in the body during hazardous situations.

LF may be influenced by the vagus nerve, sympathetic nervous system, and stress reflex mechanisms, while HF power is generated by efferent vagal nerve activity induced by respiratory activity (Shaffer et al., 2014). Therefore, a low LF/HF ratio typically

reflects more excellent activity of the PNS relative to SNS activity (Taylor, 2006). It can be observed when individuals cope with challenges requiring effort and increased sympathetic nervous system activation. In this study, the LF/HF ratio showed different values in risky situations (2, 3, and 4) compared to safe situations, generally indicating lower values in risky situations. It suggested that individuals perceived the presence of risk, activated their SNS, and exhibited lower PNS activity, keeping their bodies ready to face the environmental challenge. The magnitude of the perceived risk varies, and the extent of this challenge also differs. In relatively safe situations (1), however, the LF/HF ratio was higher, indicating a normal state of the body. However, it should be noted that the variation of this indicator did not follow a strict linear relationship with increasing levels of risk due to the complex physiological regulatory mechanisms of LF/HF. Some stress studies can also observe a similar change (Ahmed et al., 2022). In stressful environments, whether the stress originates from extreme physical conditions externally or is internally perceived, individuals instinctively adjust their physiological state to better cope with the challenges posed by stress.

(c) Analysis of nonlinear indicators

The Poincaré plot was a powerful tool for understanding HRV due to its ability to capture non-linear dynamics (Tulppo et al., 1996). While time and frequency domain indicators were commonly used to assess the ANS, these linear measures offered limited insights into HRV (Huikuri et al., 2003). The Poincaré plot represented the nonlinear dynamics of a phenomenon and unveiled hidden correlation patterns in a time series signal (Karmakar et al., 2010). It simplified consecutive heartbeat intervals into a visual representation on a cartesian plane (Brennan et al., 2001). Over time, these points formed a curve or trajectory that describes the system's evolution (Hoshi et al., 2013). The short axis of the ellipse represented the standard deviation (SD1) of instantaneous RR interval variability, while the long axis represented the standard deviation (SD2) of continuous

long-term RR interval variability. These two indicators were commonly used to assess the dynamic characteristics of HRV (Hoshi et al., 2013). SD1 was considered a parasympathetic index of sinus node control because the influence of the vagus nerve on the sinus node was quicker than sympathetic effects (Mourot, Bouhaddi, Perrey, Cappelle, et al., 2004; Mourot, Bouhaddi, Perrey, Rouillon, et al., 2004). In contrast, SD2 was influenced by parasympathetic and sympathetic tones (De Vito et al., 2002). These two indicators did not simply increase with higher levels of risk and tended to be higher during safe conditions and smaller when exposed to risk. In particular, SD1 decreased when risk was at its highest, while SD2 increased. The same pattern can be observed in individual who frequently use helmets; when riding without a helmet, they face higher risk (Fyhri & Phillips, 2013). Additionally, both indicators exhibited significant standard deviations. In conjunction with their physiological significance, it became evident that similar to the neural regulatory mechanisms revealed by time and frequency domain indicators, an enhanced sympathetic response and a weakened parasympathetic response occurred when facing risk. This trend was even more pronounced when risk was at its peak. Furthermore, the substantial standard deviations observed in each context reflected that the variability of heart rate fluctuations among different individuals was notably high.

A⁺⁺ and B⁻⁻ reflected more complex, dynamic and nonlinear changes during risky situations. The non-linear indicator A⁺⁺ had its highest value in the most dangerous situations, reflecting increased PNS activity. However, indicator B⁻⁻ also reached its highest, reflecting increased SNS activity. These two indicators lead to two contradictory results. Combining the analysis of HRV time and frequency domain indicators suggested that SNS was activated strongly and PNS activity was lower in the most dangerous situations. It implied that we could not simply consider sympathetic and parasympathetic nerves as branches having opposite effects simultaneously on the heart, meaning an increase in the activity of one does not necessarily correspond to a decrease in the other.

Current research on the physiological regulatory mechanisms of HRV has confirmed that the relationship between the two branches was complex and nonlinear. The increased PNS activity may be associated with a decrease, increase, or no change in SNS activity (Shaffer & Ginsberg, 2017). After aerobic exercise, the PNS was reactivated while the SNS remained elevated (Billman, 2013; Billman et al., 2015; Shaffer & Ginsberg, 2017). The theory known as the "free-energy principle," proposed by Friston, K. (Friston, 2009), suggested that perception, behavior, and cognitive processes can be seen as means for biological systems to reduce free energy through active inference. In specific contexts, the brain constructed models to predict external states. It then gradually adjusted its internal models based on external cues and visceral feedback until the prediction matched the actual outcome more closely. This process allowed the biological system to adapt to the environment, anticipate the future, and achieve goals more effectively. Consequently, this adjustment process was ongoing and required a continuous exchange of information between the viscera and the brain. The significant fluctuations in the A++ and B-- indicators between consecutive heartbeats represented this adjustment process in the heart organ.

In the present study, the seemingly opposite meanings of A++ and B-- indicated that both SNS and PNS activities were enhanced in risky situations, and these enhancements exhibited intense changes. While the overall trend showed a stronger SNS dominance over PNS at the micro-level of individual heartbeats, both systems exhibited rapid and drastic fluctuations. It can be called parasympathetic rebound, and it may occur following high levels of stress (Nada et al., 2001; Shaffer & Ginsberg, 2017). The reactivation and sustained activation of the PNS may be an instinct in dangerous situations, as it may indicate individuals' desire to protect themselves and escape from the hazardous environment, reflecting physiological "fight or flight" responses (Wehrwein et al., 2016).

Alternatively, it could be a natural inclination of individuals to calm themselves down to deal with critical situations, which was achieved through the strong activation of the PNS.

Furthermore, these changes reached significant levels not only in relatively safe Scenario 1 and risky scenarios 2, 3, and 4 but also across different levels of risk within each scenario. Significant differences were observed in pairs such as 2 and 3, 2 and 4. These differences were particularly pronounced in more dangerous scenarios. Moreover, even though the occurrence of danger in the risky scenarios and the occurrence of accident consequences account for only a tiny fraction of time, there were statistically significant differences in the variability of successive heartbeats during these brief periods. Clearly, the two indicators were more sensitive than HR and HF.

4.4.2.3 The changes of *PD* and *ST*

Although pupil dilation was usually suggested as an indicator of perceived risk (He et al., 2022), pupil diameter (PD) did not show significant differences in either of the scenarios. Therefore, it can play a role as supplemental information. It was reported that more than 90% of the information was received through vision (Moharreri et al., 2018) and pupil diameter showed significant changes in different conditions. For example, PD decreased as participants became more familiar with the stimulus scenarios (Zhang et al., 2022) and increased while driving with haptic feedback due to higher cognitive engagement (Pakdamanian et al., 2018). It also displayed sensitivity for measuring changes in cognitive demand in auditory–verbal–vocal tasks (Niezgoda et al., 2015). In the present experiment, PD tended to increase approximately as the level of risk rose, indicating the increased activation of SNS. The higher value observed in the riskiest situation may imply that the activities of the SNS were stronger, and at the same time, the arousal level was higher (Bradley et al., 2008). However, this change was not statistically significant. The possible reason may be that three out of the four scenarios with risks

resulted in a relatively similar cognitive load during perceiving, identifying, and assessing risk. This similarity in cognitive load led to no significant difference in PD.

Regarding the ST, it also exhibited a higher value in the last situation, indicating predominant SNS activation. However, these differences were not statistically significant. In a related experiment that investigated risk responses in construction works, a combination of various physiological signals, such as EDA, PPG, and ST, was employed. Unfortunately, this study did not analyze the changes in different physiological signals in high and low-risk situations, which hinders our understanding of whether temperature as a variable underwent significant change. The short-term measurement may not be enough for subjects to mediate their temperature.

4.5 Personal Differences

A two-level HLM is employed to analyze personal differences, which is given and explained in Section 3.8. To simplify the result, we only included indicators with personal differences.

4.5.1 Personal difference of PAD

Three dimensions of emotional reactions were analyzed using the HLM model in the driving situation, and the crucial results are listed in Table 4.8.

Table 4.7: Result of hierarchical linear modeling for PAD in driving

Variable		P		A		D	
		Model1	Model2	Model1	Model2	Model1	Model2
Step1	Risk level	0.813	0.813	-1.011	-1.011	-1.337	-1.337
Step2	Personal		-0.015		-0.005		-0.008
F		78.979**	42.892**	184.32**	74.777**	155.132**	78.079**
R^2		0.27	0.287	0.283	0.284	0.42	0.423
ΔR^2		0.27	0.018	0.283	0.001	0.42	0.003

Notes: *, $p < 0.05$, **, $p < 0.01$

The three dimensions exhibited significant differences in both risk levels and individual characteristics. Taking P as an example, a detailed explanation was given. When the first control variable, risk level, was introduced into the linear regression model (referred to as model 1), it yielded statistically significant ($F(1,216) = 78.979, P < 0.01$).

Risk level alone contributed to 27% of the variability in R^2 . Then, moving to model 2, a new personal variable, was added on the basis of model 1. The new model 2 with two independent variables was still significant ($F(2,215)=42.892, P<0.01$), and the combined contributions of risk level and personnel to the variability in R^2 increased to 28.7%, which was 1.7% higher than that of model 1. It meant the personal difference in the Pleasure dimension accounted for 1.7% of the total variances.

Regarding variable A, it exhibited a similar pattern to variable P. The "risk level" factor accounted for 28.4% of the variability in A, showing a slight increase of 0.1% compared to Model 1. On the other hand, for variable D, the contribution of "risk level" was 42%, while the personal difference accounted for 0.3%.

Based on these findings, it can be inferred that there was a genuine personal difference in emotional response, although the magnitude of this difference was relatively small.

4.5.2 Personal difference of EDA

The two indices for electrodermal activity (EDA) yielded different results. EDL showed no significant differences in either risk level (model1, $F(1, 216)=0.138, P=0.710$) or individual differences (model2, $F(2, 215)=1.898, P=0.152$). On the other hand, individual differences did contribute to the variability in EDR. In model 2 ($F(2, 215)=15.984, P<0.01$), compared to model 1 ($F(1, 216)=14.261, P<0.01$), individual differences accounted for 3.1% of the variability in EDR.

Table 4.8: Result of hierarchical linear modeling for EDR in driving

Variable		EDR	
		Model1	Model2
Step1	Risk level	0.066	0.066
Step2	Personal		-0.005
	F	14.261**	15.984**
	R^2	0.062	0.093
	ΔR^2	0.062	0.031

Notes: *, $p<0.05$, **, $p<0.01$

4.5.3 Personal difference of HRV

No significant individual differences were observed in the HRV time domain indicators analysis, except LF/HF and A++, for which model 1 showed statistical significance. However, when an individual difference variable was added in model 2, the previous statistical significance was no longer observed. Despite this, the variable contributed 0.4% to the variability in B--.

Table 4.9: Result of hierarchical linear modeling for HRV in driving

Variable		A++		B--		LF/HF	
		Model1	Model2	Model1	Model2	Model1	Model2
Step1	Risk level	0.266	0.266	0.272	0.272	-0.284	-0.284
Step2	Personal		0.002		0.006		0.014
	F	5.713*	2.881	8.044*	4.46*	4.126*	2.994
	R ²	0.026	0.026	0.036	0.04	0.019	0.027
	ΔR ²	0.026	0.00	0.036	0.004	0.019	0.008

Notes: *, $p < 0.05$, **, $p < 0.01$

4.5.4 Discussion of Personal Difference

In physiological and psychological experiments involving humans, it is typically not feasible to directly compare indicators across different stimuli in a cross-sectional manner. However, analysis of the change patterns for these indicators is still possible.

The results of the HLM analysis in this study indicated that individual factors accounted for the highest proportion of variability in the EDR indicator, followed by P. Variations in the remaining indicators were primarily attributed to different risk levels, with individual factors contributing only a minimal portion to the observed variations. The sensitivity of EDR in distinguishing individuals was consistent with previous studies. For example, novice drivers undergoing their first driving simulation training exhibited significant differences in EDA, indicating variations in arousal levels (Schnittker, 2012). Safety-oriented drivers showed higher EDA than those inclined toward risk (Liang & Lin, 2018). The EDR variations among construction workers engaged in high-risk activities also showed differences at the individual level, while EDL did not (B. Choi et al., 2019).

Regarding HRV, only B-- displayed personal differences, but the proportion was minimal. The related study analyzed that variations in HRV may be influenced by factors, such as sex, age and personality traits (Di Simplicio et al., 2012).

These results indicated that changes in physiological and emotional responses were mainly caused by different risk levels, possibly resulting from controlled subject factors. It also implied that machine learning models established on this dataset would not be influenced by personal differences and purely rely on risk levels. Participants were all in good physical health, with similar ages and shared background knowledge. These conditions can minimize the impact of individual factors on the experimental results. The results of HLM provided further evidence that the observed variations were primarily due to the risk levels.

4.6 Summary

The statistical analysis deepens our understanding of the recognition of perceived risk under different risk levels, especially under indistinct risk situations. Compared to direct measurements of self-report, the combination of multiple physiological and emotional measures, as an indirect approach, is regulated by a neuroendocrine system that modulates perceived responses to risk situations. Several conclusions can be drawn from the analysis results of the situations ranging from relatively safe to moderate risk and finally to the highest risk.

As a low-cost, simple, and feasible approach, the emotion PAD model played a significant role in the emotional assessment of risk situations. These three dimensions demonstrated a gradual transition of emotional states, becoming increasingly unhappy, with higher arousal levels and a stronger sense of losing dominance as risk increases. Especially for risk situations with similar risk levels, where negative emotions dominate, the arousal and dominance dimensions showed excellent discriminative ability.

During perceiving risk, individuals naturally experience physiological changes of varying magnitudes. Among them, the EDR component of EDA and multiple indicators in HRV, such as HR, HF, A++, and B--, showed significant differences. Remarkably, the nonlinear indicators A++ and B-- demonstrated good sensitivity even in situations with approximate risk levels. The existence and magnitude of differences in physiological data were determined by the perceived differences between the extremes of risk (Megias et al., 2018).

The relationship between neural regulatory and risk situations was explored during perceiving risk. HRV was significantly inhibited, indicating enhanced cardiac sympathetic activity and weakened vagus nerve activity. However, it was crucial to emphasize that the changes in the two branches of the ANS, which have antagonistic effects, did not strictly follow a simple and linear pattern of increase or decrease. Additionally, from the perspective of sequential heartbeats or micro-level, both the sympathetic and parasympathetic nervous systems exhibited significant increases and fluctuations in response to the presentation of danger and the consequences. This illustrated the rapid and chaotic changes happening in both branches of the human nervous system, even though the sympathetic nervous system dominates at the macro-level, especially in the highest-risk situations. These observations suggest that humans have a coping mechanism where they swiftly mobilize their internal resources to deal with the challenges of risky situations.

The main cause of significant differences in physiological and emotional responses was the factor of risk level, not personal factor. This implied that the risk classification models established on such a dataset in the next step would not be interfered by different individual factor.

Table 4.10: The main outcomes and limitations obtained from statistical analysis

No.	Main outcomes and insights	Limitations
1	When individuals are in a risky situation, they continuously gather external information and adjust their risk perception based on their personal coping abilities. As perceived risk increases, both the sympathetic and parasympathetic nervous systems become more activated, with the sympathetic nervous system showing a higher level of activity.	1) Although the risk levels of the two situations in the experimental design are quite similar, the specific threshold at which the autonomic nervous regulation mechanism fails to differentiate between them, resulting in identical emotional and physiological responses, remains undetermined in the current study. 2) The autonomic nervous regulation mechanism is highly complex and may vary with changes in the risk situation. The duration required for these changes to be reflected in emotional and physiological signals is another issue that this study could not address.
2	When individuals face risky situation, they emotions tend to become more negative, aroused and they may lose control of the scenario. Since the situations involve risk, negative emotions are predominant. The levels of arousal and control can indicate the perceived magnitude of the risk	
3	The changes of autonomic nervous system can also be reflected in various physiological indicators simultaneously. Among these indicators, different measures vary in their sensitivity to the complex changes. Specifically, electrodermal activity and heart rate variability (both instantaneous and frequency-domain indicators) are particularly sensitive and can effectively capture this intricate internal neural dynamic.	

In summary, the statistical analysis has provided us with a deeper understanding of the physiological and psychological responses elicited in risk situations, particularly at closely comparable risk situations. It was precisely the existence of these differences that laid the foundation for the subsequent establishment of risk classification models.

CHAPTER 5: MACHINE LEARNING FOR PERCEIVED RISK ASSESSMENT

5.1 Introduction

Understanding the physiological and psychological changes in risk situations marks the initial stage of risk perception research. Building upon this groundwork, it is crucial to use them as input variables to delve deeper into the concealed correlation between the numerical changes in these indicators and risk levels, and to establish models fitting this relationship.

Previous studies have explored this domain. However, the developed models were typical binary classifiers, able only to categorize risk situations as either low or high, overlooking scenarios where risk ratings were close. Furthermore, these models solely utilized physiological signals as input vectors, neglecting the inclusion of emotionally generated responses in emergencies (B. G. Lee et al., 2021). The two aspects are the focal points of this study.

5.2 Feature Importance

Twenty variables were used as input data to train ML models. If there were correlations among these indicators, the performance of ML algorithms could be adversely affected. Therefore, it's necessary to detect and eliminate multicollinearity. The calculation of feature importance can provide a promising method to address this issue.

5.2.1 Correlation of input variables

The heatmap of the correlation matrix of twenty input variables is shown in Figure 5.1. There were two areas with significant correlation coefficients. The first one involved three dimensions of emotions, and the second included specific indicators of HRV. It indicated the presence of multicollinearity among these 20 variables.

Before determining the importance of these indicators, measures should be taken to eliminate the impact of multicollinearity. Recursive Feature Elimination (RFE) can be

used to select indicators or features. Through an optimization process, RFE can mitigate the effects of multicollinearity, thereby obtaining the best combination of indicators.

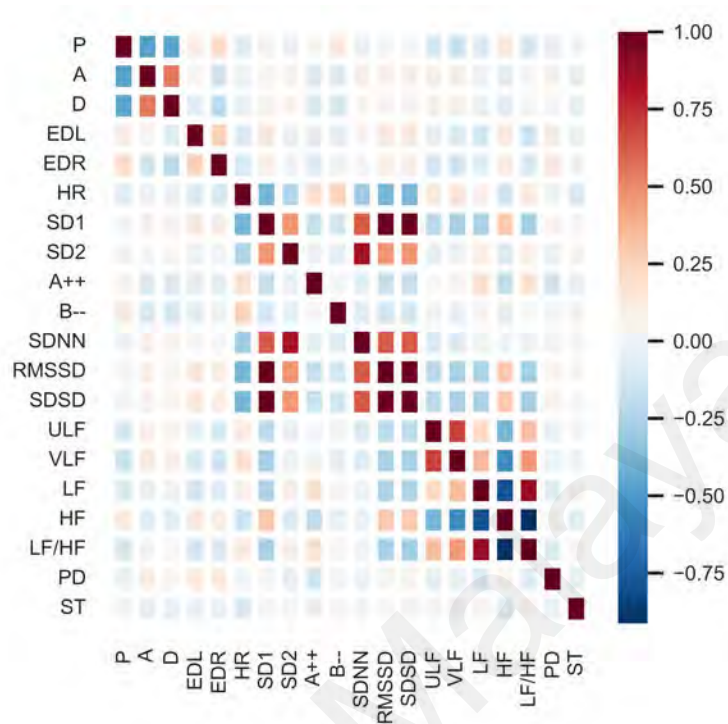


Figure 5.1: The heatmap of twenty input features in driving

5.2.2 Recursive Feature Elimination (RFE) to select variables

RFE was employed to select variables, and the trained RF was utilized as the base model. The process of RFE was distributed as follows.

- i. Import the relevant third-party libraries.
- ii. Import and scale the dataset using the StandardScaler method to compress them.
- iii. Instantiate the RF model and set it as the standard model.
- iv. Conduct a 5-fold cross-validation approach to progressively calculate the importance of indicators using RFE based on accuracy.
- v. Retrieve the column names of the optimal features and visualize the changes in cross-validation accuracy with the number of indicators.

Finally, it was discovered from Figure 5.2 that RF achieved the highest accuracy when retaining eight features. These features are P, A, D, EDR, A++, B--, ULF, and VLF.

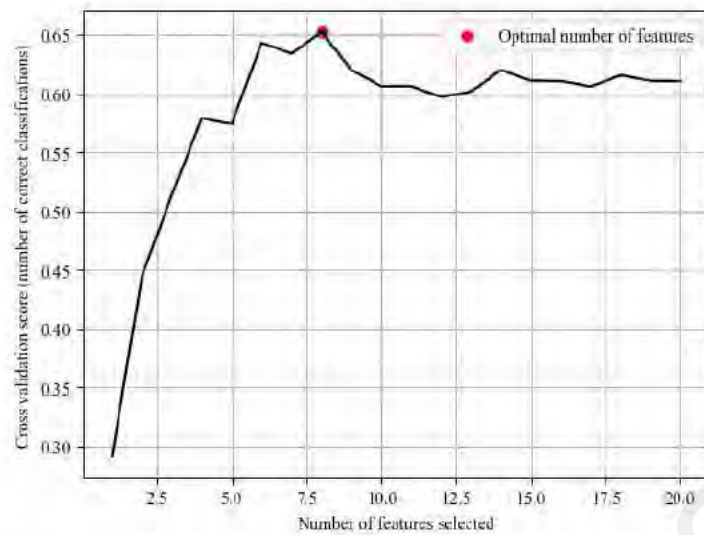


Figure 5.2: Feature selection results of RFE cross-validation in driving

5.2.3 Calculation of Feature importance

Gini importance and Permutation Importance (PI) were used to represent feature importance. Especially, Figures 5.3 and 5.4, respectively, displayed the results of them.

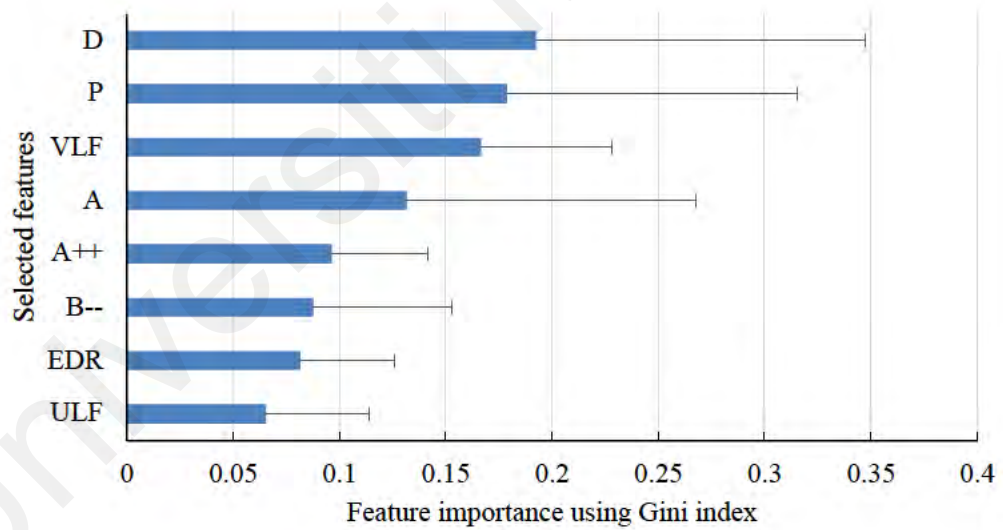


Figure 5.3: Feature importance of different indicators in driving using Gini

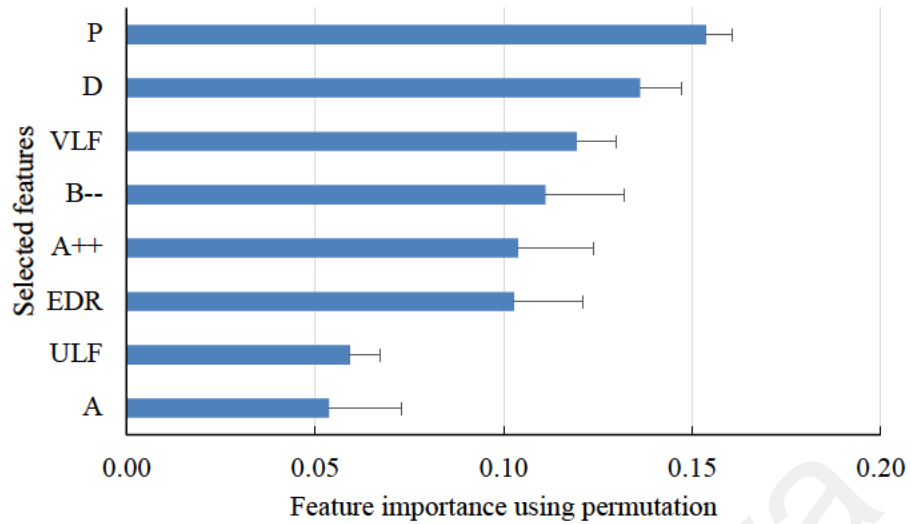


Figure 5.4: Feature importance of different indicators in driving using permutation

We categorized the results into three groups based on the range of feature importance values to compare the results obtained from both methods. These categories were subsequently arranged in descending order and are displayed in Table 5.1. Regarding the two distinct importance metrics, the rankings of the eight features showed some differences. Notably, P and D exhibited significant variation, while the rankings of the remaining seven features were remarkably similar.

Table 5.1: The results of grouping the feature importance based on Gini and Permutation methods in driving

Method	Data Range	Features
Gini	>0.15	D, P, VLF
	0.10-0.15	A
	<0.10	A++, B--, EDR, ULF
Permutation	>0.15	P
	0.10-0.15	D, VLF, B--, A++, EDR
	<0.10	ULF, A

Obviously, the left eight features can be utilized as input data to develop machine learning models.

5.3 Perceived Risk Classification

The implementation process for three fundamental machine learning models is illustrated in Figure 5.5. Although the fundamental process remained the same, variations

existed in specific parameters and hyperparameters among the three algorithms. For example, in RF, common parameters include the number of estimators, minimum samples of leaf and minimum samples of split. SVC involved parameters such as the kernel function and the values of degree and gamma. Typically, the value of C served as its hyperparameter. Regarding ANN, parameters included the number of layers, dropout layers and the activation function, with the number of epochs serving as its hyperparameter.

The implementation process is detailed below:

- i. Import necessary third-party libraries, including commonly used ones: numpy, pandas, sklearn.preprocessing, sklearn.model selection, and tensorflow.keras.

- ii. Eliminate features with higher collinearity, then load the dataset with less features and classification labels.

- iii. Define a function to scale the data using the minmaxScaler method, compressing it within the range of -1 to 1.

- iv. Split the dataset into training, validation, and testing sets based on two scenarios.

The splitting ratios are 0.7:0.15:0.15 and 0.5:0.25:0.25, respectively.

- v. Define parameters and hyperparameters, along with their respective variation ranges for different models.

- vi. Create and initialize the machine learning models (RF, SVC, or ANN).

- vii. Train the models using the training dataset. The iteration criterion was the negative classification accuracy, aiming to discover optimal parameter combination for achieving the highest classification accuracy. Hyperopt was employed for automatic hyperparameter optimization.

- viii. Utilize the obtained optimal parameters to calculate diverse evaluation metrics for the model, such as the confusion matrix and classification report, the false positive rate

(FPR) and true positive rate (TPR) for a specific class in the ROC curve, and the area under the ROC curve (AUC) for the training, validation, and testing sets.

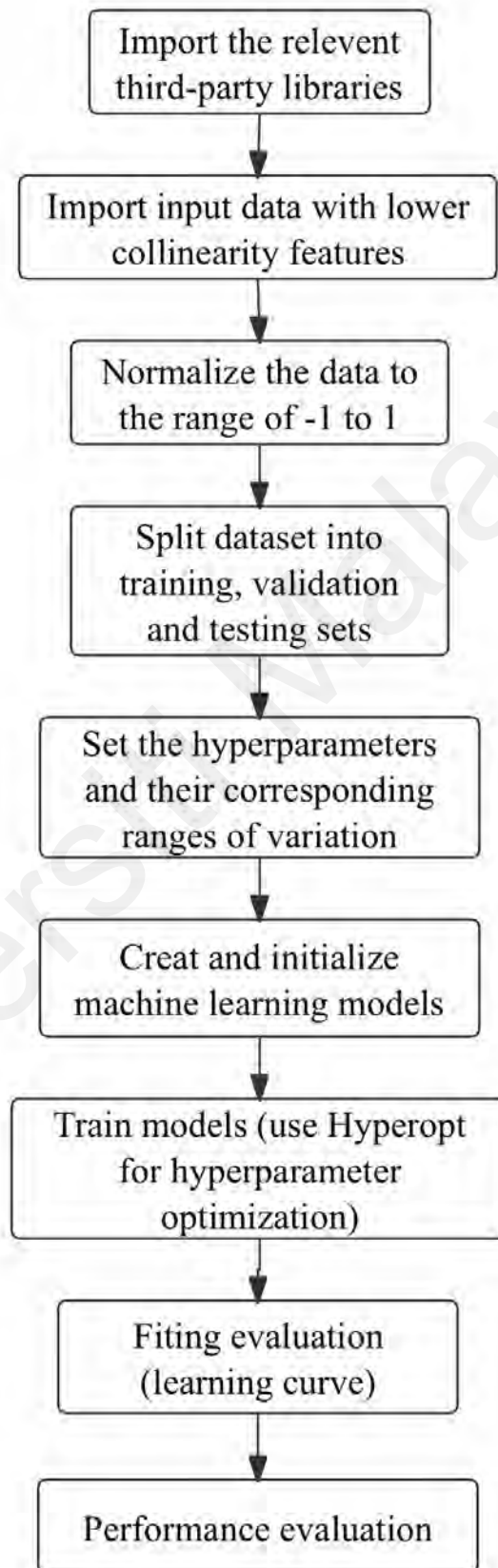


Figure 5.5: The implementation process for three basic machine learning models

5.4 Results of Machine Learning

5.4.1 Fitting Evaluation

5.4.1.1 Fitting performance of RF

Fitting performance is the prerequisite condition for machine learning. Both underfitting and overfitting are not suitable for the next prediction. Fifty-four subjects were obtained during the driving experiment. There were four situations, so there were 216 samples. Two RF learning curves for the first dataset, one with a 0.7:0.15:0.15 ratio split and the other with a 0.5:0.25:0.25 ratio split for training, validation and test datasets, are shown in Figures 5.6 and 5.7.

The two curves demonstrated that as the number of samples increased, the accuracy of both the train and validation datasets improved gradually. Furthermore, the divergence between the train and validation datasets diminished in the case of the first splitting ratio. However, for the second splitting ratio, the accuracies appear less stable.

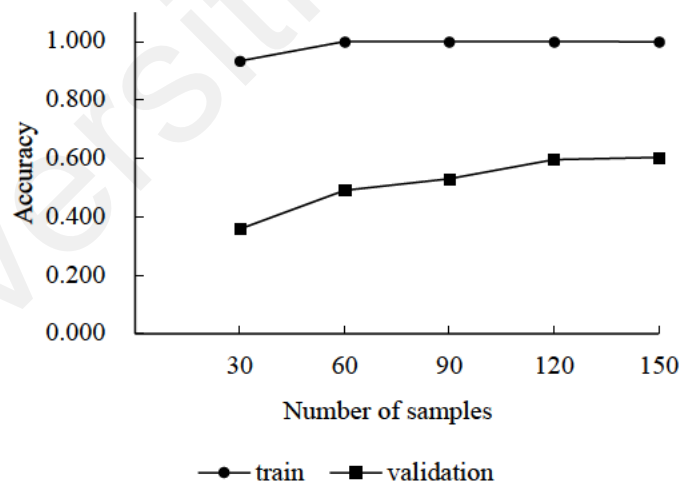


Figure 5.6: The learning curve of RF with a 0.7:0.15:0.15 ratio split for training, validation, and test datasets in driving

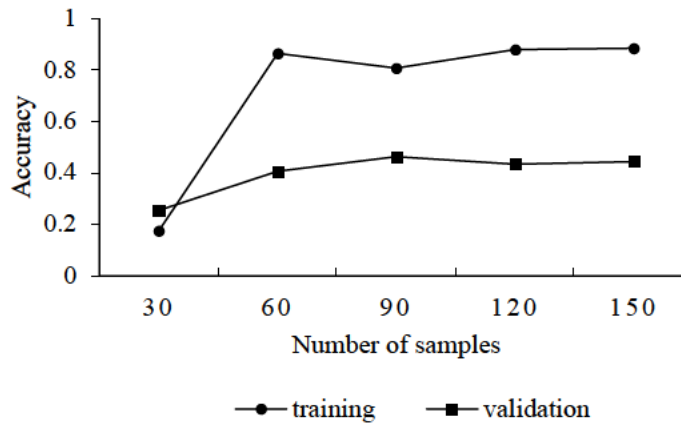


Figure 5.7: The learning curve of RF with a 0.5:0.25:0.25 ratio split for training, validation, and test datasets in driving

The optimized parameters of RF were listed in table 5.2.

Table 5.2: The optimized parameters of RF

No.	parameters	values of optimized parameters	
		0.7:0.15:0.15	0.5:0.25:0.25
1	n_estimators	245	225
2	min_samples_split	3	4
3	min_samples_leaf	3	2
4	bootstrap	False	False

5.4.1.2 Fitting performance of SVC

The learning process of SVC was similar to RF. Figures 5.8 and 5.9 show that increasing the number of training samples, correlates with the improved accuracy of both the training and validation datasets. Moreover, it was evident that with 70% of data used for training, the difference in accuracy between the training and validation datasets became smaller.

It is worth noting that accuracy values with a 50% training dataset were higher than those with a 70% training dataset. The differences in accuracy between the training and validation data were more significant with a 0.5:0.25:0.25 split than with a 0.7:0.15:0.15 split.

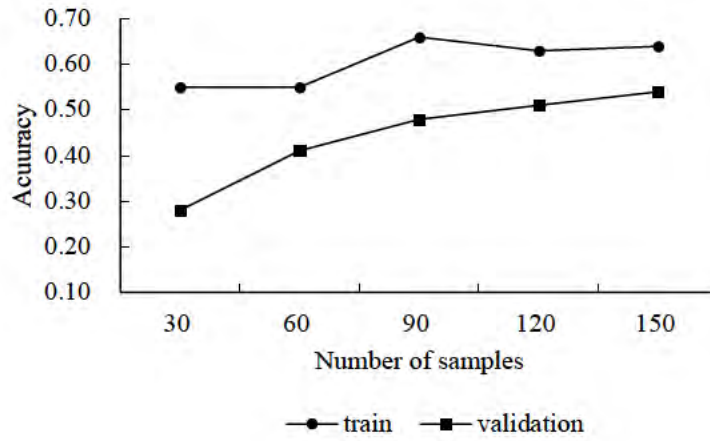


Figure 5.8: The learning curve of SVC for driving experiment data with a 0.7:0.15:0.15 ratio split for training, validation, and test datasets in driving

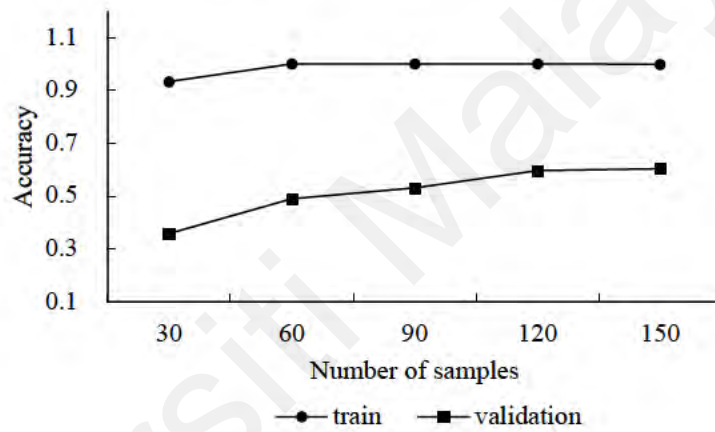


Figure 5.9: The learning curve of SVC for driving experiment data with a 0.5:0.25:0.25 ratio split for training, validation, and test datasets in driving

The optimized parameters of SVC were listed in table 5.3.

Table 5.3: The optimized parameters of SVC

No.	parameters	values of optimized parameters	
		0.7:0.15:0.15	0.5:0.25:0.25
1	C	4.06	5.32
2	degree	7.00	7.00
3	gamma	0.03	0.04
4	kernel	sigmoid	sigmoid

5.4.1.3 Fitting performance of ANN

The learning curve of the ANN utilized the Root Mean Square Deviation (RMSD) to measure the fitting quality. Observing Figures 5.10 and 5.11. It is evident that, with less than 50% of the training data, the RMSD of both the train and validation datasets decreased steadily and synchronously. However, with 70% of the training data, although the RMSD also decreased, the reduction in the validation dataset was minor compared to the former ratio. It indicated that, unlike RF and SVC, the distribution of the data significantly affected the fitting results of the ANN.

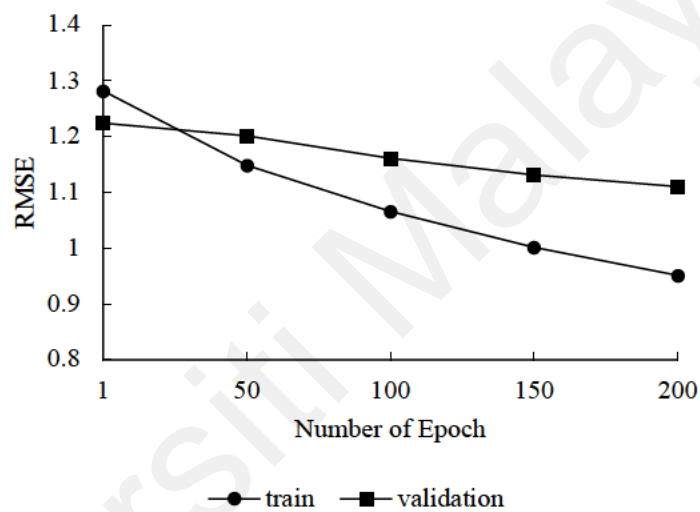


Figure 5.10: The learning curve of ANN for driving experiment data with a 0.7:0.15:0.15 ratio split for training, validation, and test datasets in driving

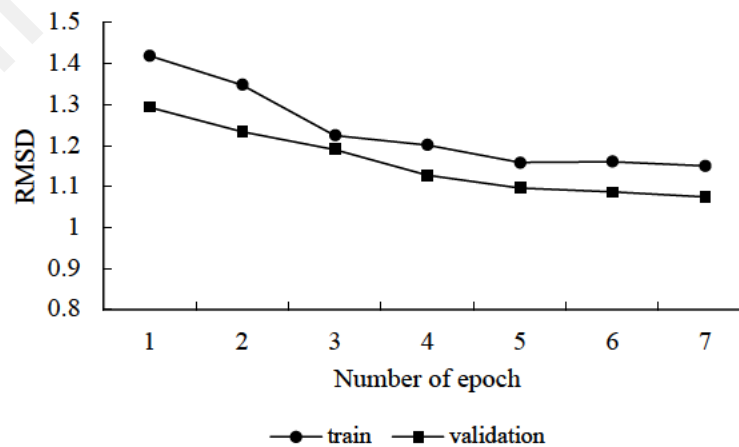


Figure 5.11: The learning curve of ANN for driving experiment data with a 0.5:0.25:0.25 ratio split for training, validation, and test datasets in driving

The optimized parameters of ANN were listed in table 5.4.

Table 5.4: The optimized parameters of ANN

No.	parameters	values of optimized parameters	
		0.7:0.15:0.15	0.5:0.25:0.25
1	number of layers	3	3
2	number of neurons	15	17
3	activation function	relu	relu
4	dropout layers	1	1

The learning curves of RF, SVC, and ANN demonstrated no underfitting or overfitting, indicating that the models fit the data appropriately.

5.4.2 Based Model Performance

The test dataset results were conducted and tested to evaluate the models' performances. Figure 5.12 displayed the confusion matrix and result reports, while Figure 5.13 showed the ROC curve and the corresponding AUC values. It is important to note that both figures presented the same testing results in different formats: data and curves.

Figure 5.12 displayed that the ANN mode, trained with 50% of the data, outperformed both RF and SVC models in precision, accuracy and recall. Moreover, the classification performance metrics of the ANN consistently surpassed those of RF and SVC across all four risk situations. Regardless of the model employed, the classification performance was optimal for the highest-risk level. Situation 4 had the highest risk score, followed by situation 1. In contrast, situations 2 and 3 exhibited closely aligned risk scores, resulting in a relatively lower performance.

F1 scores for different models and classes were employed to illustrate the findings. For instance, the F1 score for situation 4 reached 0.71, while the F1 score for risk level 1 was 0.63. Specifically, for classes 2 and 3 in RF, with a model trained using 70% of the data, the F1 scores were only 0.62 and 0.35, respectively. Simultaneously, SVC demonstrated similar changes in F1 score, with the only difference in the numerical values. On the contrary, the ANN model, trained with 50% of the data, achieved the highest F1 score across all risk levels. Specifically, this model excelled in classifying situations 4 and 1,

achieving F1 scores of 0.87 and 0.85, respectively. The significance lay in its robust performance in the more challenging situations 2 and 3, achieving F1 scores of 0.81 and 0.61, respectively.

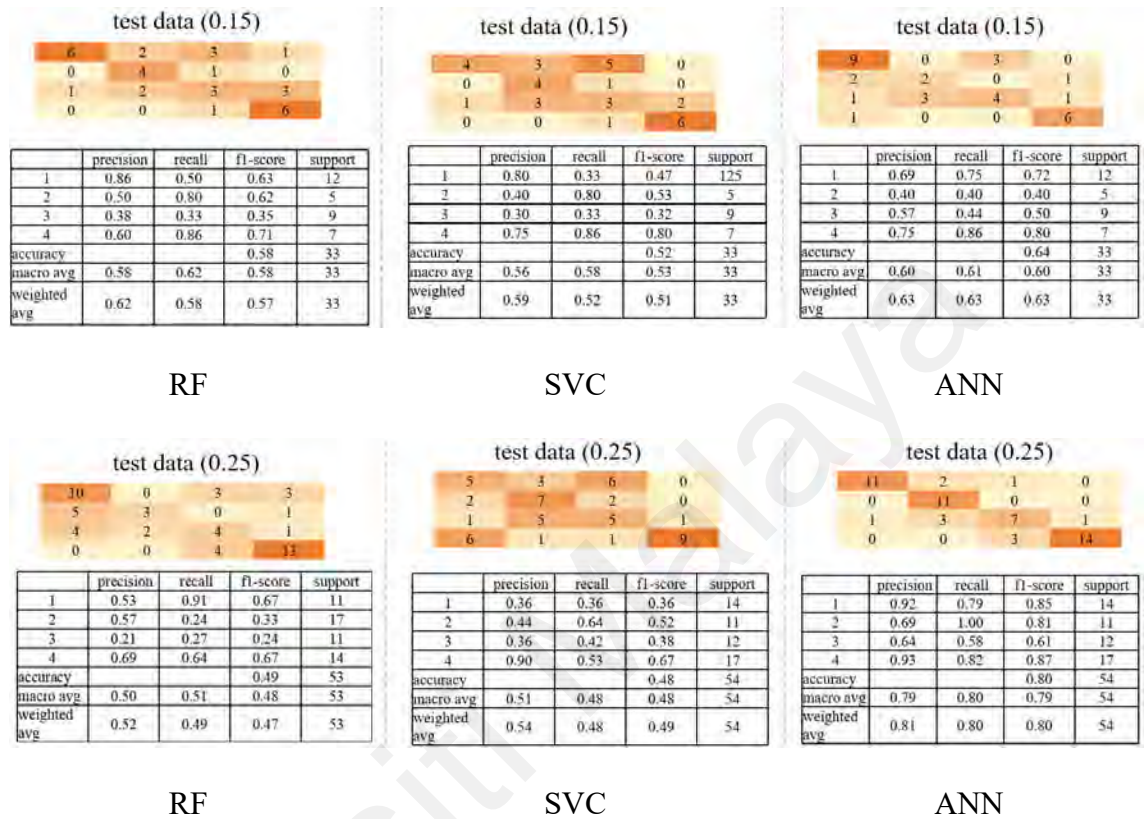


Figure 5.12: Confusion matrix and result reports of three models with two ratios split for training, validation, and test datasets in driving

The ROC curves of the three models are shown in Figure 5.13. It was observed that class 4 consistently appeared closer to the top-left corner, irrespective of the model and splitting ratio. It indicated that class 4 exhibited the highest accuracy and the lowest classification difficulty. Conversely, class 1 was inferior to Class 4 in accuracy and classification difficulty. Furthermore, class 2 and class 3 consistently presented higher classification difficulty and demonstrated lower accuracy than class 1 and class 4. They posed an increased challenge across all three models. Notably, even in the ANN model trained with 70% of the data, the ability to classify class 2 and class 3 was nearly equivalent to random guessing. Thirdly, when considering the AUC values, the ANN

model trained with 50% of the data outperformed the other models, achieving an AUC of 0.90, indicating superior performance.

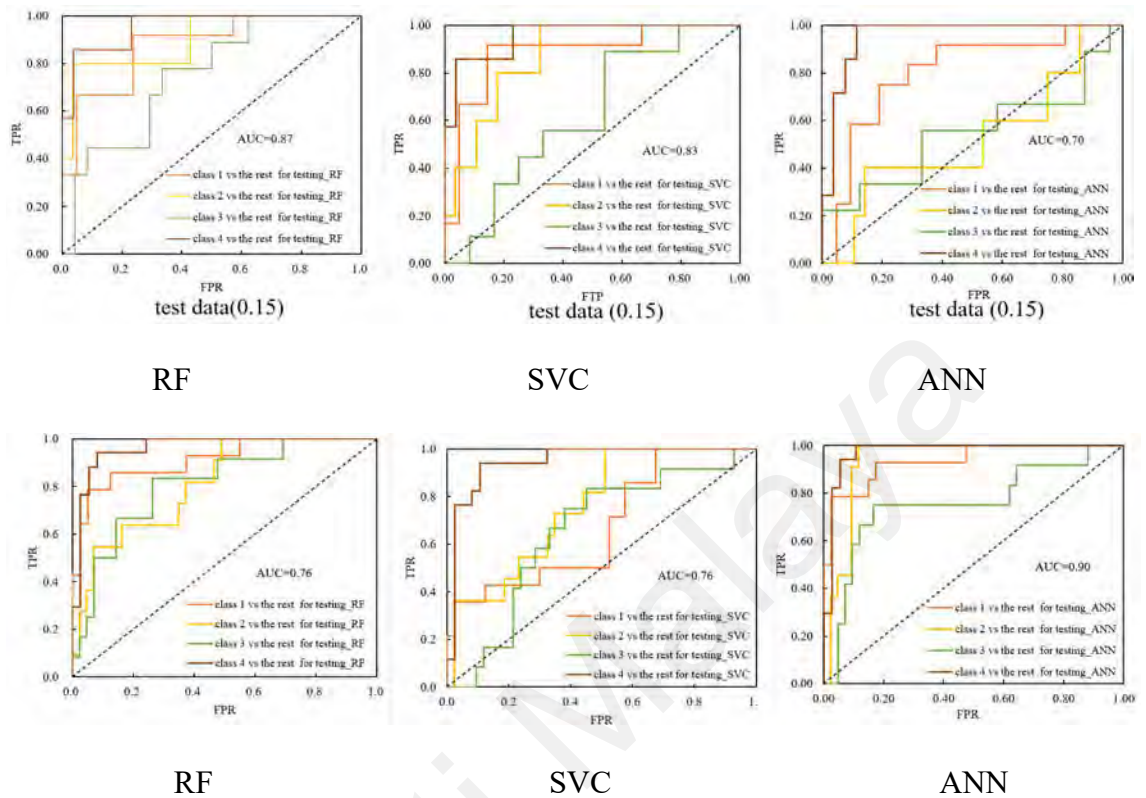


Figure 5.13: The ROC curves and AUC values of three models with two ratios split for training, validation, and test datasets in driving

5.4.3 Integrated Models Performance

In Section 3.9.2, two approaches for integrating three trained models were discussed. The first involved the computation of weight coefficients for different models and categories. The second employed the stacking technique to introduce an additional RF model on top of the existing three models, and this RF model was trained using the outputs of the initial three models. The results of the test data for both approaches are illustrated in Figure 5.14.

From the perspective of the confusion matrix and result reports of the integrated model, both approaches showed improvements in the model's performance across each classification. When compared to the best-performing model of ANN, the weight coefficient integration method increased the accuracy from 0.80 to 0.88, while the

stacking technique integration method raised it to 0.85. Notably, for the challenging-to-distinguish classes 2 and 3, the F1 scores improved from ANN's 0.81 and 0.61 to 0.91 and 0.75 (for the weight coefficient integration model), respectively, and to 0.73 and 0.80 (for the stacking technique integration model). However, there was a recurring issue in both ANN and the two integration models where the precision values for classes 2 and 3 were higher than the recall values. The combination of high precision and low recall indicated that the models were somewhat cautious when identifying positive instances, ensuring their predictions of positive instances were almost always correct. However, it may need to include some positive samples.

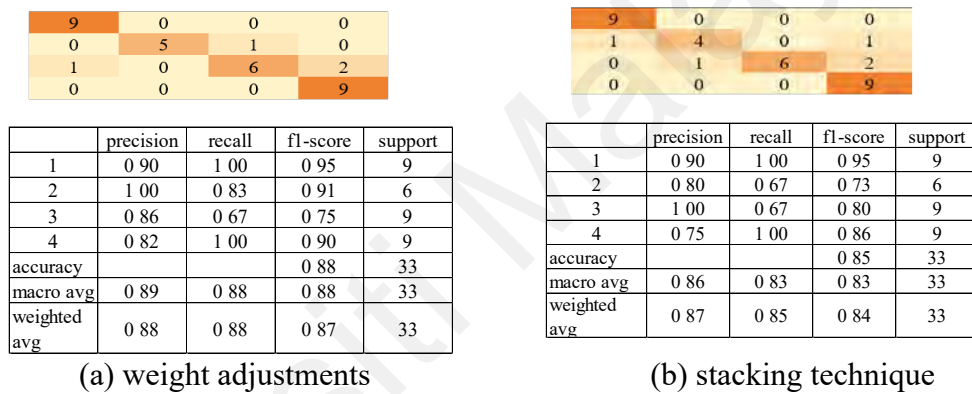


Figure 5.14: Confusion matrix and result reports of integrated model

When adjusting model output using the weight coefficient computation method, it was only possible to calculate the confusion matrix and results, such as accuracy and precision for the test data. However, with the utilization of the stacking technique, a new RF model was introduced into the training process, allowing for the computation of ROC curves and AUC values for the test data. This outcome was illustrated in Figure 5.15, where the AUC value was 0.98, surpassing the AUC of ANN, which was 0.90.

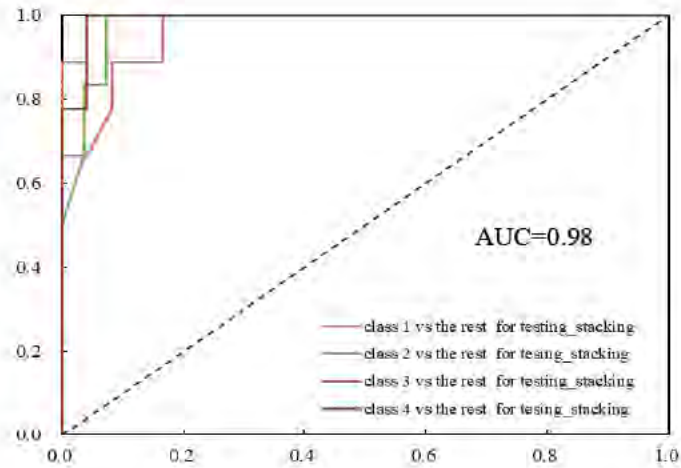


Figure 5.15: The ROC curves and AUC values of the integrated model with stacking technique

5.5 Discussion

Features lacking or minor statistical significance and high-related indicators with HRV were removed through feature selection using RFE. For instance, EDL was a typical indicator showing minor differences, and PD and ST displayed no significant differences. Besides, some features exhibited significant differences, each of which was related to HRV and demonstrated collinearity with other HRV indicators, such as HR, HF, and RMSSD.

The emotion dimensions P and D exhibited relatively higher importance scores. However, the importance of three dimensions should be paid more attention. Regarding dimension A, it occupied a mid-level position according to the Gini index but had the smallest value under the Permutation index. Both A and D displayed a negative correlation with P and a positive correlation with each other. Consequently, when using permutation to calculate feature importance, removing or shuffling either A or D had an equal impact on the classification accuracy of RF. Thus, A demonstrated the lowest importance in the context of feature importance as represented by Permutation index. Conversely, in the Gini index, the importance of A and D was closely comparable. The

results highlighted the significance of emotional reactions induced by risk situations as crucial internal risk perception sensitivity indicators.

Physiological indicators representing short-term heart rate variations demonstrated remarkable performance distinguishing risk levels. For example, the instantaneous heart rate nonlinearity indices A++ and B-- and the HRV frequency domain features VLF, ULF, and EDR showed notable importance. The previous studies (B. G. Lee et al., 2021) have compared EDA, PPG, and the combination of all data (EDA, PPG, and ST) to demonstrate the importance of EDA and PPG in predicting subjective risk. At the same time, it did not further specify which indicators within EDA and PPG were important. However, this study utilized two methods for calculating feature importance (Gini index and Permutation index) after adopting RFE to eliminate collinearity, yielding relatively consistent and comprehensive results.

RF, SVC, ANN and two integrated machine learning models were used to classify risk levels using eight features. The input vectors included physiological signals and emotional reactions. Furthermore, this study not only considered low and high-risk situations. It involved situations where the risk degrees were relatively close. The close degrees significantly increased the classification difficulty for machine learning.

ANN surpassed RF and SVC for four risk levels. Its F1 score achieved 0.85 and 0.87 for low and high-risk scenarios; for risk levels 2 and 3, the F1 score was 0.81 and 0.61, respectively. The overall accuracy for the four classes reached 0.80. Notably, all these indices to evaluate the classification performance of ANN were higher than those of RF and SVC, demonstrating the superior nonlinear fitting capability of ANN over them. It should be emphasized that these results were obtained under the precondition that ANN utilized training sets with similar distributions across different classifications. Compared to specifically related studies (B. G. Lee et al., 2021), this research developed a binary classification model to differentiate low (relatively safe) and high-risk (dangerous)

scenarios. However, the input vectors used were different. This study contained a combination of on-site collected worker bio-signals, including EDA, ST, and PPG. Moreover, the F1 score of the optimal Gaussian SVM model achieved 0.82 for low and 0.81 for high-risk level, which was smaller than the scores of 0.85 and 0.87 obtained by ANN for low and high-risk scenarios, respectively.

The performance of integrated models exceeded that of ANN, especially in situations with subtle differences in risk degree. Regardless of the weight adjustment or stacking technique models, they had higher accuracies, precisions and F1 scores than ANN. For example, the F1 score of situations 2 and 3 reached 0.91 and 0.75 in the weight-adjusting model and 0.73 and 0.80 in the stacking model. However, the values of the best single model, ANN, were 0.81 and 0.67. For low and high-risk levels (1 and 4), F1 was 0.95 and 0.90 for the weight adjustment model, 0.90 and 0.86 for the stacking model, whereas the F1 score of ANN was 0.85 and 0.87.

5.6 Summary

The main findings and results are summarized in Table 5.5.

Table 5.5: The main outcomes and limitations obtained from modeling

No.	Main outcomes and insights	Limitations
1	When classifying perceived risk based on emotional and physiological changes, indicators with larger ranges of variation and greater independence from other metrics contribute more significantly to the classification.	Although the risk levels of the two scenarios in the experimental design are quite similar, the specific threshold at which simple machine learning models lose their ability to classify them remains undetermined in this study.
2	The stronger the emotional and physiological changes induced by the risk scenario, the easier and more accurate the classification becomes. Simple models are sufficient to meet classification requirements in these cases.	
3	When individuals perceive only minor differences in risk levels between scenarios, classification becomes much more challenging. Conventional models may fail to differentiate effectively, necessitating the use of ensemble techniques to enhance classification performance.	

Features that exhibit significant differences and do not correlate strongly with other indicators are more likely to be retained. Additionally, the PAD emotion model can

effectively represent emotional reactions induced by risk scenarios and exhibits high feature importance in the classification model. Additionally, indicators representing instantaneous heart rate variability demonstrate elevated feature importance. Moreover, two feature importance calculating approaches yield similar results, albeit in different ways.

For any machine learning model, the riskiest scenario demonstrates the highest accuracy and the slightest difficulty in classifying among the four situations. The relatively safer situation ranks below the highest risk in accuracy and classification difficulty. However, situations that fall into the intermediate risk level present lower accuracy and more incredible classification difficulty.

Integrated models demonstrate a superior capability in distinguishing risks with relatively subtle differences in risk degrees. This classification proficiency further substantiates its potential applications, feasibility and generalization of machine learning algorithms in risk perception. Specifically, for situations with substantial differences in risk magnitude, ANN suffice. However, in cases where the disparities in risk are minimal, the use of ANN alone fails to yield satisfactory classification results. It is where integrated models come into play. In other words, without the necessity of employing more intricate deep learning algorithms, a straightforward approach of model integration using weighted methods over the base models exhibits discriminative power, allowing the capture of subtle physiological and emotional distinctions induced by minor variations in risk levels.

CHAPTER 6: VALIDATION FOR GENERALIZED CAPABILITY OF MACHINE LEARNING

6.1 Introduction

The application of machine learning technology for assessing perceived risk has yielded significant results, particularly in situations characterized by a comparable degree of risk. The effective utilization of physiological and emotional responses played a crucial and positive role in this context. This chapter aimed to elucidate the application paradigm of this assessment method. Additionally, a new construction scenario was introduced as a case study to validate the generalized ability of this evaluation mode.

6.2 The Application Paradigm of Risk Perception Assessment

The method of assessing risk perception has been successfully applied in a driving context to categorize perceived risk, utilizing physiological and emotional changes as input data. This process differs from the traditional engineering risk methods. It is essential to extract the generic paradigm for risk assessment. Furthermore, to validate the paradigm, an alternative application context is proposed. Previous studies have primarily centered on construction sites, designating it as the second context. The application paradigm for evaluating perceived risk will be summarized. Subsequently, the new trial serves as a case study to confirm the proposed assessment paradigm.

The application paradigm for assessing perceived risk through physiological and emotional responses is displayed in Figure 6.1.

- i. Determine risk situations. Situations characterized by a high probability, or severe consequences can be determined as investigation objectives. Workers in such contexts may exhibit deficiencies in risk perception, highlighting the need to identify hazard factors and then take measures to reduce risk. These situations could be visually presented in video to illustrate the entire process.

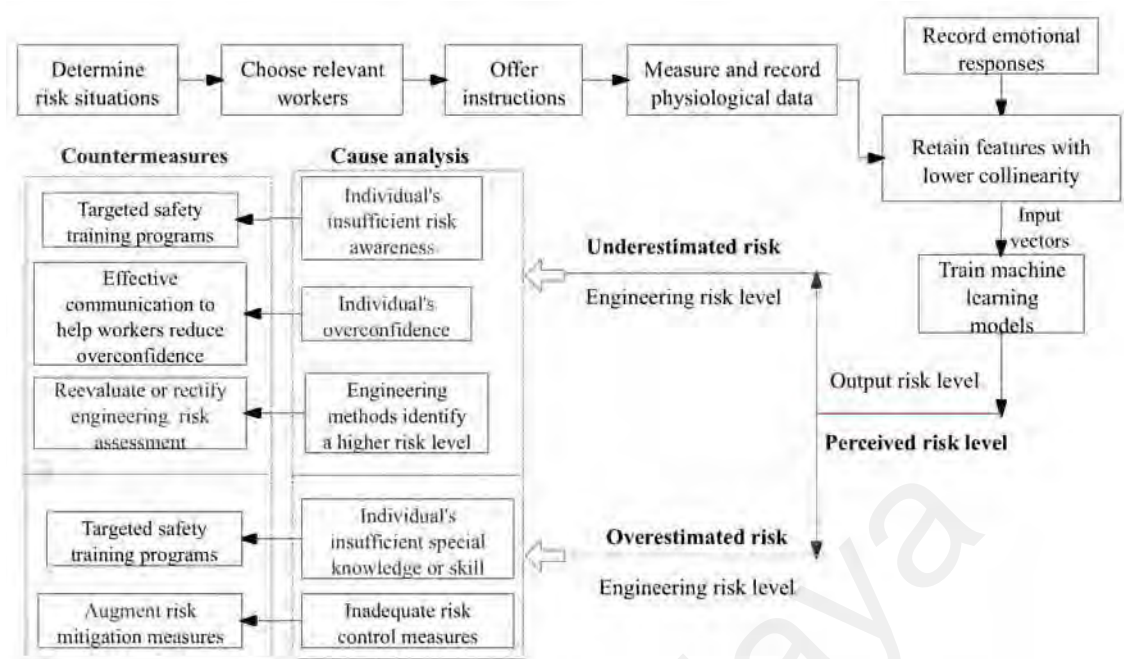


Figure 6.1: The application paradigm for assessing perceived risk through physiological and emotional responses

ii. Choose relevant workers. Personnel required to operate in risk situations could serve as experimental subjects, especially in high-risk work environments or positions, where all available candidates are considered relevant personnel.

iii. Offer instructions. Instructions should be provided before the entire measurement process. Some key points, such as relaxation before measuring and maintenance focus during the measurement procedure, should be emphasized to the participants.

iv. Measure and record physiological data. Workers wear physiological sensors. These responses can be measured and recorded synchronously while workers observe risk situations. Then, the raw data can be preprocessed by ErogLAB and physiological indicators can also be calculated through ErogLAB software.

v. Record emotional responses. SAM is utilized to obtain values from three emotional dimensions, helping workers express their emotional reactions.

vi. Eliminate features that exhibit high collinearity with other indicators, naturally retaining a small number of indicators that carry important information as input data.

vii. Train machine learning models. The combination of physiological and emotional indicators is split into training, validating and testing data. The training data are used as input vectors to train three standard and two integrated models. The validation data are used to assess fitting performance. The prediction of test data, including the confusion matrix and result reports, is used to compare the performance of different models.

viii. Compare perceived risk with engineering-assessed risk.

If the perceived risk level is lower than the engineering-assessed risk, there is a risk of underestimation. Possible reasons may include individual insufficient risk awareness or overconfidence. Another possibility is that engineering methods identify a higher risk level. Enhancing individuals' safety awareness through personalized safety training, reducing overconfidence through effective communication, and reassessing or correcting existing engineering risk assessment results are all part of implementing successful safety measures.

If the perceived risk level is higher than the engineering-assessed risk, there is a risk of overestimation. Possible causes may include an individual's insufficient specialized knowledge or skills and inadequate risk control measures. Similarly, safety managers should provide additional safety training and education to enhance their cognition and understanding of the specific risk, potentially reducing their perceived risk to a normal level. Alternatively, safety professionals need to reanalyze current management and control actions and implement additional technical measures to mitigate the risk.

In summary, this comparison can assist safety managers in identifying individuals with flawed risk perceptions or situations with potential hazards. Additionally, the final assessment result may need to be calibrated based on a combination of engineering assessment results and workers' risk perception.

6.3 Validation for Statistical Analysis Results

The validation experiment was carried out in accordance with the procedures outlined in Figure 3.1. Statistical and machine learning methods were employed to analyze the physiological and emotional data. The findings from the construction scenario will be compared and discussed alongside the results from the driving context.

6.3.1 Validating PAD Analysis Results

60 subjects were recruited in this experiment, but 7 lost their physiological data. So, there were 53 groups of data P, A and D would be analyzed. Similarly, the results for each situation are listed in the Table 6.1.

Table 6.1: Non-parameter test and pair comparison results of P, A, and D in construction

Variables	P				A				D			
Group	1	2	3	4	1	2	3	4	1	2	3	4
Media	6	7	7	8	6	5	4	2	6	6	5	3
25%	5	5	6	8	5	3	3	1	4	4	2	1
75%	7	7	8	9	7	6	4	3	7	7	6	4
H	67.839**				90.033**				40.871**			
Cohen's f	0.075				0.09				0.061			
Pair comparison (Mann-Whitney U test)	test	Cohen's d			test	Cohen's d			test	Cohen's d		
	1-4, t=346.5**	1.65			1-4, t=2626*	2.33			1-4, t=2301.5**	1.32		
	2-4, t=323**	1.62			2-4, t=2456.5**	1.66			2-4, t=2228*	1.18		
	3-4, t=561.5**	1.16			3-4, t=2331.5**	1.32			3-4, t=1892.5*	0.66		
	1-3, t=925.5*	0.66			1-3, t=2246.5**	1.14			1-3, t=1828*	0.56		
	2-3, t=1050*	0.46			1-2, t=1944*	0.65			2-3, t=1747 P=0.059			
	1-2, t=1243.5 P=0.612				2-3, t=1773.5*	0.45			1-2, t=1495.5 P=1.126			

Note: *, $p < 0.05$, **, $p < 0.01$

To better understand how the three dimensions evolve with changing risk levels, the values of the three dimensions are shown in Figure 6.2. The patterns of change in the three dimensions were the same as those observed in the driving scenario: P ranged from slightly unhappy (6) to very unhappy (8), A transitioned from a low arousal state (6) to a highly aroused state (3), while D exhibited a decreasing trend, indicating an increased sense of losing control.

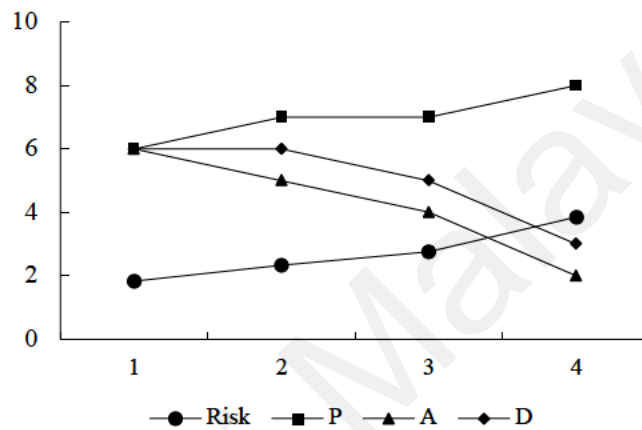


Figure 6.2: The changing trend of P, A, and D with risk ranks in construction

The differences in all three dimensions were statistically significant in the Kruskal-Wallis test. Notably, the effect size for dimension A (0.09) was higher than that of P (0.075) and D (0.061). It contrasted with the findings from the driving experiment, where D exhibited a higher effect size than P and A.

Regarding the pairwise comparison results, the underlying patterns remained consistent despite inconsistencies between the two scenarios. Specifically, P, A, and D exhibited greater magnitudes between safety (1) and risk situations (2, 3, 4). However, when the risk levels were closely matched, such as between 2 and 3 or 3 and 4, the values of Cohen's *d* were smaller. It was worth noting that Cohen's *d* values differed from those observed in the driving experiment. When the difference in risk levels was slight, A had the most incredible discriminative ability, followed by P, with D having only discriminative ability. However, A still exhibited statistically significant differences in

scenarios where the risk levels were very close (1-2 and 2-3). When comparing 1 and 2, the effect size was 0.65, indicating a medium-level difference. Similarly, the effect size was 0.45 when comparing 2 and 3, indicating a subtle difference. However, dimension P showed differences between 2-3 with an effect size of 0.46, while dimension D did not show significant differences in either of these scenarios.

6.3.2 Validating Physiological Analysis Results

6.3.2.1 Statistical analysis of EDA

The EDL data from Table 6.2 exhibited minor variations within the range of 7 to 8.5, with the highest value observed in the fourth situation, which had the highest risk level. The data for this indicator approximately followed a normal distribution but did not meet the assumption of sphericity ($\chi^2 = 65.414$, $p < 0.001$). After applying the Greenhouse-Geisser correction, significant differences were found ($F(1.661, 86.386) = 15.584$, $p < 0.001$). Similarly, the EDR data displayed a similar pattern of changes, with the highest value observed in the fourth situation. However, this indicator approximately followed a normal distribution and met the assumption of sphericity. Repeated-measure ANOVA revealed significant differences as well.

When comparing the effect sizes, it was found that EDL ($\eta^2 = 0.231$) showed a significant effect size than EDR ($\eta^2 = 0.188$), which differed from the findings in driving. Considering post hoc comparisons, EDL only showed significant differences in driving between situations 3 and 4. However, it exhibited significant differences between situations 4 and 1, as well as 4 and 2. In contrast, the post hoc comparisons for EDR were similar to those observed in driving, indicating substantial differences between situations 4 and 1, as well as 4 and 2. EDR still demonstrated significant differences between situations 1 and 3, particularly notable in the closely related risk levels between situations 2 and 3. Furthermore, when assessing the magnitude of these differences using Cohen's d

values, the highest values were observed in the comparison between situations 1 and 4, regardless of whether it was EDL (-0.325) or EDR (-0.679).

Table 6.2: Description and repeated-measure ANOVA results of EDA in construction

Variables	EDL				EDR			
Group	1	2	3	4	1	2	3	4
Mean	7.01	7.49	7.40	8.50	0.22	0.23	0.33	0.38
Std.	4.31	4.47	4.58	4.97	0.20	0.19	0.27	0.29
Test of Sphericity	$\chi^2 = 65.414, P < 0.001$				$\chi^2 = 7.221, P = 0.205$			
Within Subjects Effects	F (1.661, 86.386) = 15.584, p < 0.001				F (3, 156) = 12.010, p < 0.001			
η^2	0.231				0.188			
Post Hoc Comparisons	T test		Cohens'd		T test		Cohens'd	
	1-4, -1.49**		-0.325		1-3, -0.11*		-0.439	
	2-4, -1.01**		-0.220		1-4, -0.16**		-0.679	
	3-4, -1.10**		-0.240		2-3, -0.09*		-0.383	
					2-4, -0.15**		-0.625	

Notes: *, $p < 0.05$, **, $p < 0.01$

6.3.2.2 Statistical analysis of HRV

The four time-domain indicators of HRV did not show a sustained trend of decline or increase; the third and fourth situations primarily influenced their fluctuations. Additionally, most indicators reached their minimum values in the highest-risk situation. Furthermore, except for the HR indicator (F (1.935, 100.620) = 12.690, $p < 0.001$), statistical analysis did not reveal significant differences among the other indicators. These findings were consistent with those observed in the driving scenario.

When conducting post hoc comparisons of HR, we observed significant differences between the safe situations 1 and 2 and between 1 and 4. Additionally, there were significant differences observed between situations 1 and 3. Consequently, the effect size of HR in the construction scenario ($\eta^2 = 0.196$) was higher than that observed in the driving ($\eta^2 = 0.102$).

Table 6.3: Description and repeated-measure ANOVA results of HRV time-domain indicators in construction

Variables	HR/bpm				SDNN/ms			
Group	1	2	3	4	1	2	3	4
Mean	78.19	75.74	75.51	76.40	65.39	66.06	81.48	65.70

Table 6.3, continued

Variables	HR/bpm				SDNN/ms			
Group	1	2	3	4	1	2	3	4
Std.	10.18	9.96	10.59	9.94	52.48	51.57	119.16	48.12
Test of Sphericity	$\chi^2 = 42.087, P < 0.001$				$\chi^2 = 147.722, P < 0.001$			
Within Subjects Effects	F (1.935, 100.620) = 12.690, p < 0.001				F (1.217, 63.293) = 1.047, p = 0.325			
η^2	0.196							
Post Hoc Comparisons	T test		Cohens'd					
	1-2, 2.45**		0.241					
	1-3, 2.68**		0.263					
	1-4, 1.79*		0.176					
Variables	RMSSD/ms				SDSD/ms			
Group	1	2	3	4	1	2	3	4
Mean	69.30	63.21	90.59	63.57	69.54	63.40	90.32	63.71
Std.	76.27	74.63	166.11	71.19	76.53	74.85	166.72	71.36
Test of Sphericity	$\chi^2 = 143.017, P < 0.001$				$\chi^2 = 143.173, P < 0.001$			
Within Subjects Effects	F (1.237, 64.341) = 1.407, p = 0.246				F (1.237, 64.320) = 1.405, p = 0.246			

Notes: *, $p < 0.05$, **, $p < 0.01$

The frequency indicators of HRV, descriptions, and statistical results are presented in Table 6.4. The five indicators were analyzed using the same analytical approach as in the driving experiment. To facilitate the observation of trends, the values of these indicators are plotted in Figure 6.3.

Table 6.4: Description and repeated-measure ANOVA results of HRV frequency-domain indicators in construction

Variables	ULF				VLF			
Group	1	2	3	4	1	2	3	4
Mean	2.14	2.43	2.23	2.84	24.42	26.37	24.63	28.36
Std.	1.40	1.46	1.42	1.56	11.55	12.79	11.96	11.69
Test of Sphericity	$\chi^2 = 4.452, P = 0.486$				$\chi^2 = 2.465, P = 0.782$			
Within Subjects Effects	F (3,156) = 3.584, p = 0.015				F (3,156) = 2.024, p = 0.113			
η^2	0.064							
Post Hoc Comparisons	T test		Cohens'd					
	1-4, -0.69*		-0.474					
Variables	LF				HF			
Group	1	2	3	4	1	2	3	4
Mean	37.79	39.29	37.17	38.31	35.65	31.92	35.98	30.50
Std.	14.92	13.08	14.12	11.04	18.81	16.83	17.09	14.21
Test of Sphericity	$\chi^2 = 5.327, P = 0.377$				$\chi^2 = 4.130, P = 0.531$			
Test of Sphericity	$\chi^2 = 5.327, P = 0.377$				$\chi^2 = 4.130, P = 0.531$			
Within Subjects Effects	F (3,156) = 0.558, p = 0.644				F (3,156) = 3.776, p = 0.012			
η^2					0.068			
Post Hoc Comparisons					T test		Cohens'd	
					3-4, 5.48*		0.326	
Variables	LF/HF							
Group	1	2	3	4				
Mean	1.96	1.98	1.55	1.65				
Std.	2.80	1.98	1.44	1.06				
Test of Sphericity	$\chi^2 = 53.894, P < 0.001$							
Within Subjects Effects	F (1.886, 98.062) = 1.495, p = 0.230							

Notes: *, $p < 0.05$, **, $p < 0.01$

Similar to the driving situations, the LF and HF components dominated the frequency domain of HRV, accounting for 30%-40% of the total. However, there was a difference in the second-highest proportion, where the VLF component accounted for 20%-30%, higher than in the first experiment. In contrast to the first experiment, where ULF, VLF,

and HF displayed significant differences, only ULF and HF showed significant statistical differences in the second experiment.

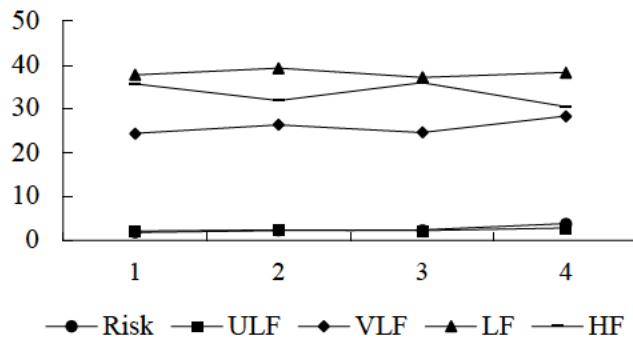


Figure 6.3: Mean values of HRV frequency-domain indicators with risk level in construction

When considering the nonlinear indicators of HRV in comparison to the driving scenarios, SD1 remained consistent across all scenarios, while SD2 displayed diverse changes without a continuous and consistent pattern. The highest values for SD1 and SD2 (63.87 and 93.46, respectively) were observed in the third situation, while the lowest values (44.83 and 76.23, respectively) occurred in the second risky and first safer situations.

As for A++ and B--, they exhibited a clear upward trend as the risk level increased, with significant differences noted, accompanied by effect sizes of 0.606 and 0.570, respectively. The most significant difference was observed between the first and fourth situations, with Cohen's d values of -1.282 and -1.150, respectively. The difference decreased between the first and third situations (-0.936 for A++ and -0.753 for B--), followed by the second and fourth situations with lower degrees of difference. The degrees of difference for the remaining paired groups were all below -0.5.

Table 6.5: Description and repeated-measure ANOVA analysis results of HRV nonlinear indicators in construction

Variables	SD1/ms				SD2/ms			
Group	1	2	3	4	1	2	3	4
Mean	49.17	44.83	63.87	45.05	76.23	79.92	93.46	79.44
Std.	54.12	52.92	117.88	50.46	53.93	53.41	112.38	48.83
Test of Sphericity	$\chi^2 = 143.174, P < 0.001$				$\chi^2 = 145.245, p < 0.001$			
Within Subjects Effects	F (1.237,64.320) = 1.405, p = 0.246				F (1.226,63.770) = 0.925, p = 0.359			
Variables	A++				B--			
Group	1	2	3	4	1	2	3	4
Mean	38.42	43.98	50.88	55.49	30.74	36.76	40.26	45.28
Std.	11.09	11.81	14.49	15.41	11.22	12.06	12.84	14.27
Test of Sphericity	$\chi^2 = 12.757, P = 0.026$				$\chi^2 = 16.674, P = 0.005$			
Within Subjects Effects	F (2.625,136.501) = 79.934, p < 0.001				F (2.556,132.895) = 69.062, p < 0.001			
η^2	0.606				0.570			
Post Hoc Comparisons	T test		Cohens'd		T test		Cohens'd	
	1-2, -5.57**		-0.418		1-2, -6.02**		-0.476	
	1-3, -12.47**		-0.936		1-3, -9.53**		-0.753	
	1-4, -17.08**		-1.282		1-4, -14.55**		-1.150	
	T test		Cohens'd		T test		Cohens'd	
	2-3, -6.91**		-0.518		2-3, -3.51**		-0.277	
	2-4, -11.51**		-0.864		2-4, -8.53**		-0.674	
	3-4, -4.60**		-0.346		3-4, -5.02**		-0.397	

Notes: *, $p < 0.05$, **, $p < 0.01$

6.3.2.3 Statistical of PD and ST

The variations in both the PD and ST indicators were relatively small. PD showed a gradual increase with the rise in risk level. ST also exhibited its highest value in the riskiest situations. These results were consistent with the findings of the first experiment.

Furthermore, the data for both indicators satisfied the normal distribution assumptions and variance homogeneity. Hence, an ANOVA test was conducted, revealing no significant differences in either indicator across the different risk levels. This consistency with the findings from the driving suggested similar patterns in the present study.

Table 6.6: Description and repeated-measure ANOVA analysis results of PD and ST in construction

Variables	PD				ST			
	1	2	3	4	1	2	3	4
Group								
Mean	3.71	3.73	3.84	3.89	33.55	33.73	33.66	34.90
Std.	1.11	1.07	1.23	1.15	2.80	3.76	3.37	3.51
Test of Sphericity	$\chi^2 = 1.584, P=0.903$				$\chi^2 = 1.972, P=0.853$			
Within Subjects Effects	F (3,156) = 0.326, p=0.806				F (3,156) = 1.872, p=0.137			

6.3.3 Discussion and Comparison with Previous Statistical Findings

6.3.3.1 Discussion and comparison with emotional responses

(a) *Changes in emotional state*

The temporal evolution of emotional states paralleled driving scenarios, transitioning from calmness in safety to a gradual shift towards negativity, arousal, and a sense of losing control as risk escalates. Moreover, this composite emotional intensity becomes more pronounced at the peak of risk.

(b) *Differentiation ability among the three dimensions*

The three emotional dimensions also showed significant differences across four levels of risk in construction scenarios, and the magnitude of differences was larger in risk levels with significant disparities and lower in situations with closer risk proximity. These findings were the same as the driving scenario. However, their differentiation abilities varied according to Cohen's *f*. D exhibited the highest value, followed by A and P in the driving context. Conversely, A demonstrates better discriminatory ability, followed by P and D in the confirmation experiment. This inconsistency may be attributed to the driving and construction scenarios. In driving, situations 2 and 3 represented unsuccessful accidents, while in the construction scenario, both situations involved hazards and casualties, triggering unpleasant emotions. However, the level of arousal varied significantly based on the degree of danger and the consequences of the accidents. In simple terms, the ability of dimension P to distinguish negative emotions decreased,

which aligned with previous research findings on the relationship between emotions and purchasing behavior. It suggested that the ability of P to express subtle differences in an unhappy environment is limited.

The discriminative ability of D in construction was reduced compared to driving. When comparing the risk disparity between the driving and construction scenarios, the gap in risk between 1 and 4 narrowed, and the risk scores between 2 and 3 were also closer.

6.3.3.2 Discussion and comparison with physiological responses

While the primary changing patterns on construction sites paralleled those seen in driving, including similar shifts in emotional states and a more vital ability to differentiate risk levels based on significantly different indicators, there were also changes in the activity of the SNS. At the same time, the PNS decreased as risk rose at a macro-level perspective, with both exhibiting extreme oscillations at the micro-level. Despite these parallels, there were still specific distinct differences. In this context, we will further explore these disparities.

Looking at EDA as a starting point, EDL and EDR displayed significant differences when exposed to driving and constructional stimuli. Notably, EDL showed a greater degree of differentiation in the latter scenario. This difference may be attributed to the risk level of the final clip in the driving scenario, which was more dangerous than the risk situations depicted in the construction site scene. It potentially triggered an instantaneous and higher identifiable arousal response due to a more potent stimulus (Horn et al., 2020), resulting in a significant change in EDR. Conversely, in the construction scene, all three video clips (2, 3, and 4) depicted accident consequences, leading to a slower and overall higher level of tonic arousal, which induced a greater EDL (Boucsein et al., 2012).

In terms of HRV, only ULF and HF under risk construction situations showed significant statistical differences. This disparity could be attributed to the relatively

similar levels of risk (2.55, 2.69 for risk situations 2, 3) observed compared to the driving scenarios (2.45, 2.85 for the close two risk situations). The average value of risk in construction was higher than driving. This factor resulted in A++ and B-- displaying greater fluctuation and oscillation and showing significant differences when comparing within-group variances.

Consistently, the results for PD and ST showed no significant differences. It further supported the notion that cognitive load similarity existed across different risk situations for PD, and short-term measurement did not provide enough time for neural mediation of body temperature.

6.3.4 Confirmation of Personal Difference Results

6.3.4.1 Personal difference of PAD

Similar to the driving scenario, all three dimensions (P, A, and D) showed significant individual differences. The contribution of individual differences to the variability of P, A, and D was found to be 4.6%, 0.9%, and 1.4%, respectively. Moreover, it was observed that P had the highest contribution, while A had the lowest. These findings aligned with the driving situation.

Table 6.7: Result of hierarchical linear modeling for PAD in construction

Variable		P		A		D	
		Model1	Model2	Model1	Model2	Model1	Model2
Step1	Risk level	0.809	0.809	-1.154	-1.154	-0.885	-0.885
Step2	Personal		-0.024		0.013		-0.018
	F	79.104**	49.091**	144.59**	74.777**	46.772**	25.46**
	R^2	0.274	0.32	0.408	0.417	0.182	0.196
	ΔR^2	0.274	0.046	0.408	0.009	0.182	0.014

Notes: *, $p < 0.05$, **, $p < 0.01$

6.3.4.2 Personal difference of EDA

EDL, as a fundamental and tonic component of EDA, displayed no significant difference in risk level ($F(1, 212) = 2.435$, $P = 0.120$) and also showed no variation related to personal factor ($F(2, 211) = 1.831$, $P = 0.163$). In contrast, EDR presented individual differences when comparing model 2 ($F(2, 211) = 8.79$, $P < 0.001$) and model 1 ($F(1, 212)$)

=15.75, $P < 0.001$), with personal factor contributing only 0.8%. These findings were consistent with those observed in the driving scenarios, albeit with different contribution percentages for EDR.

Table 6.8: Result of hierarchical linear modeling for EDR in construction

Variable		EDR	
		Model1	Model2
Step1	Risk level	.058	0.058
Step2	Personal		0.001
	F	15.75**	8.79**
	R^2	0.07	0.078
	ΔR^2	0.07	0.008

Notes: *, $p < 0.05$, **, $p < 0.01$

6.3.4.3 Personal difference of HRV

LF/HF showed differences only on risk level in the driving experiment but significantly differed on personal factors. Individuals can explain a 2.4% variation for LF/HF. Both A++ and B-- had personal differences, which accounted for 1.7% and 8% for the variation of A++ and B--.

Table 6.9: Result of hierarchical linear modeling for HRV in construction

Variable		LF/HF		A++		B--	
		Model1	Model2	Model1	Model2	Model1	Model2
Step1	Risk level	-0.135	-0.135	5.81	5.81	4.72	4.72
Step2	Personal		0.02		0.12		0.26
	F	1.30	3.24*	50.88**	28.037**	37.09**	31.67**
	R^2	0.006	0.030	0.195	0.212	0.15	0.23
	ΔR^2	0.006	0.024	0.195	0.017	0.15	0.08

Notes: *, $p < 0.05$, **, $p < 0.01$

6.3.4.4 Discussion and Comparison with Personal Difference

The main findings were consistent with those observed in driving. For instance, the same indicators of three dimensions of emotion (P, A and D) and EDR, B++ demonstrated personal differences. Furthermore, their contributions for variable variations were also relatively lower when compared to the impact of the risk factor. However, A-- displayed individual differences under the latter stimuli. It may be attributed to the higher averaged

risk scores, which aligned with EDL having a higher effect size and A++ and B-- showing more significant oscillations.

6.4 Validation for Machine Learning Performance

The process of calculating feature importance and training three machine learning models of RF, SVC and ANN remained consistent with the previous experiment, including the integrated models. However, the data were altered to reflect different situations, transitioning from driving to construction. Similarly, data were split into training, validation, and test sets using two techniques, with ratios of 0.7:0.15:0.15 and 0.5:0.25:0.25, respectively, as previously employed.

6.4.1 Comparison of Feature Importance

6.4.1.1 Correlation of input variables

The correlation heatmap of different input vectors is shown in Figure 6.4. Similarly, three emotional dimensions of P, A and D, and several indicators of HRV, such as RMSSD and SDSD, SD1, SD2 and SDNN, HF and LF/HF, EDR and EDL, had higher correlation relationships. These indicated that the collinear issue of input vectors should be considered before conducting feature importance.

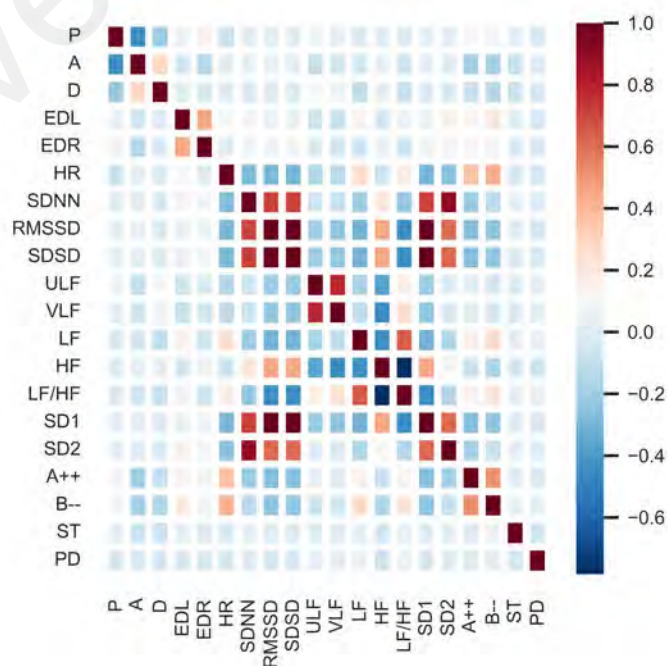


Figure 6.4: The heatmap of twenty input features in construction

6.4.1.2 RFE to select variables

The combination of RF and RFE was still employed to eliminate of redundant features. As shown in Figure 6.5, ten indicator combinations had been identified as the optimal subset of features, including A, D, EDR, HR, A++, B--, LF/HF, LF, SD2, and VLF, which was more than the driving situations. Additionally, the accuracy decreased as the number of features continually increased.

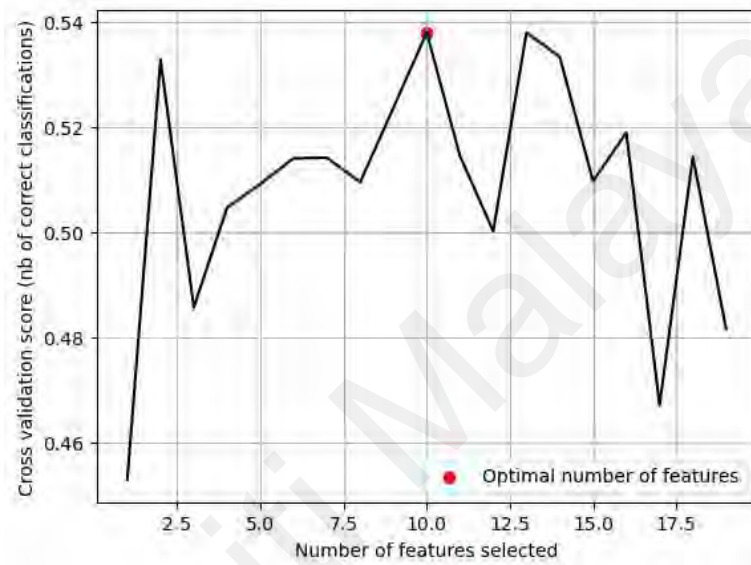


Figure 6.5: Feature selection results of RFE cross-validation in construction

6.4.1.3 Feature importance

When calculating the importance of selected features using the Gini and Permutation index, the ranking of the top ten features can be observed from Figures 6.6 and 6.7, and they were not entirely identical but primarily differed by one or two positions. The best five indicators were consistent, involving two emotional dimensions, A and D, and three HRV indicators, A++, B--, and HR.

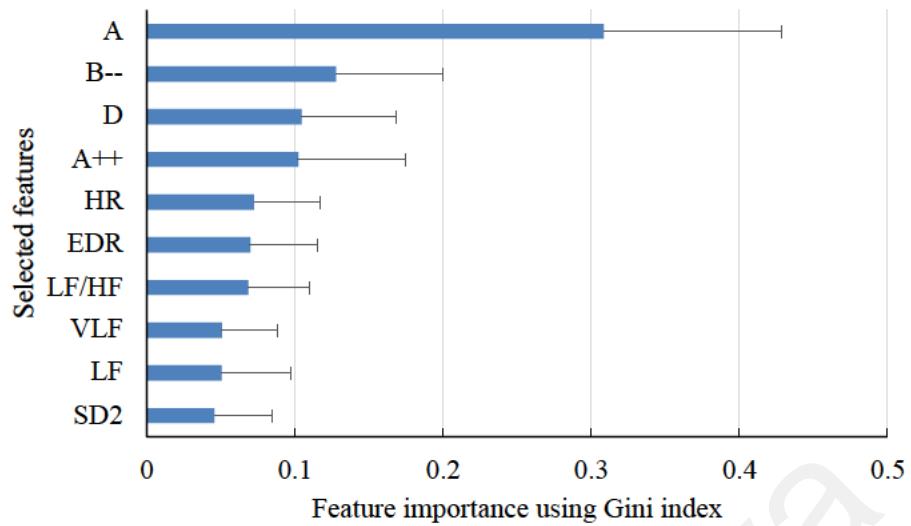


Figure 6.6: Feature importance of different indicators in construction using Gini

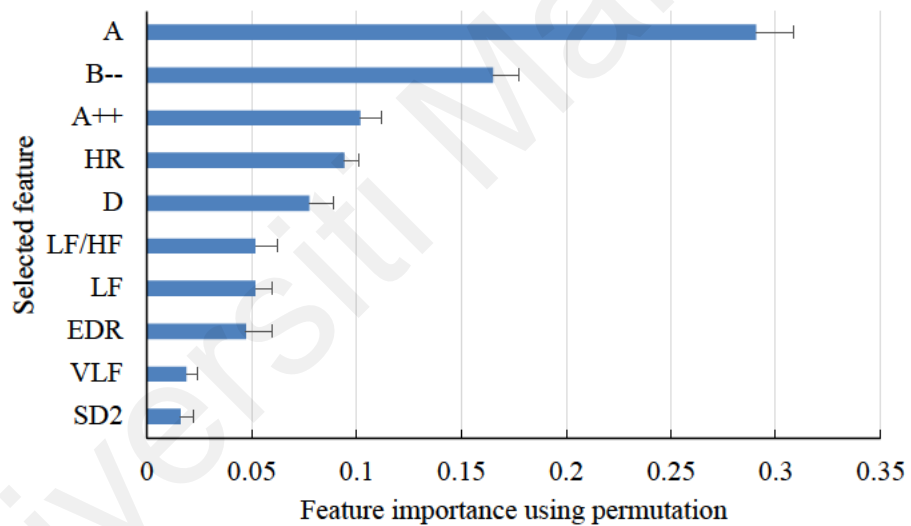


Figure 6.7: Feature importance of different indicators in constructing using permutation

6.4.2 Fitting Evaluation

Learning curves for RF, SVC and ANN trained using two different split ratios (0.7:0.15:0.15 and 0.5:0.25:0.25) were plotted between Figure 6.8 and 6.13. In addition to the ANN model trained with a 0.7:0.15:0.15 ratio, all other models exhibited typical learning curves for the training and validation datasets. It meant that as the amount of training and validation data increased, the accuracy of the models (RF and SVC) steadily increased, and the ANN model's RMSD gradually decreased, ultimately stabilizing at a

relatively steady level. However, it was worth noting that training the ANN model with 70% of the data might have been affected by an imbalanced distribution of data among the four categories, which impacted the stability of the ANN model.

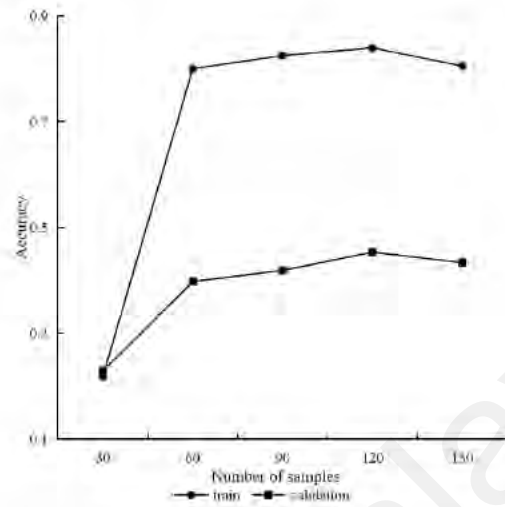


Figure 6.8: The learning curve of RF with a 0.7:0.15:0.15 ratio split for training, validation, and test datasets in construction

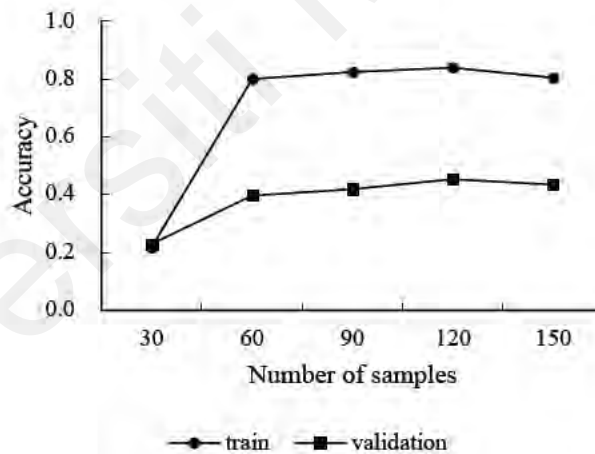


Figure 6.9: The learning curve of RF with a 0.5:0.25:0.25 ratio split for training, validation, and test datasets in construction

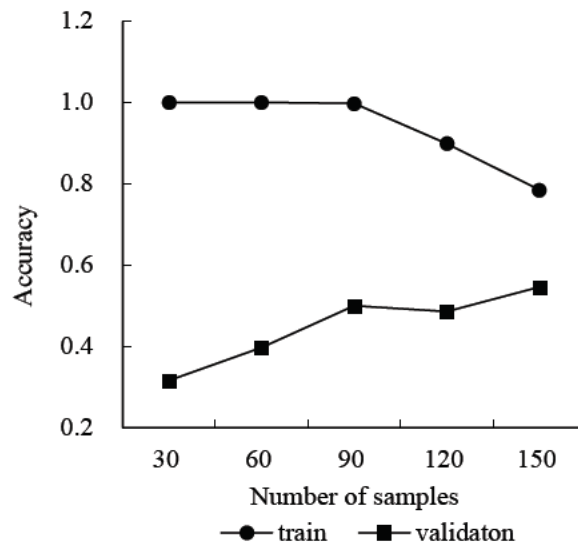


Figure 6.10: The learning curve of SVC with a 0.7:0.15:0.15 ratio split for training, validation, and test datasets in construction

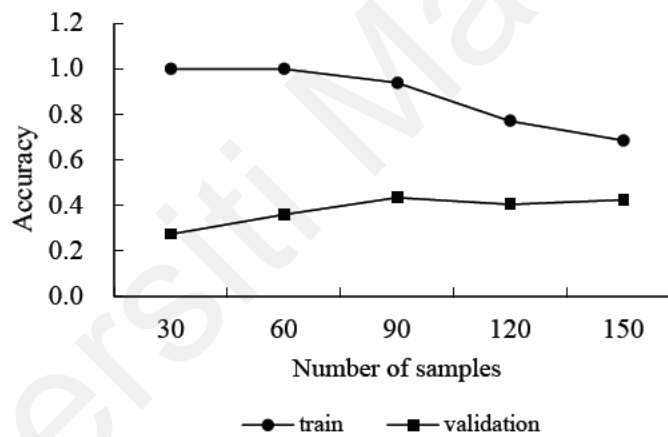


Figure 6.11: The learning curve of SVC with a 0.5:0.25:0.25 ratio split for training, validation, and test datasets in construction

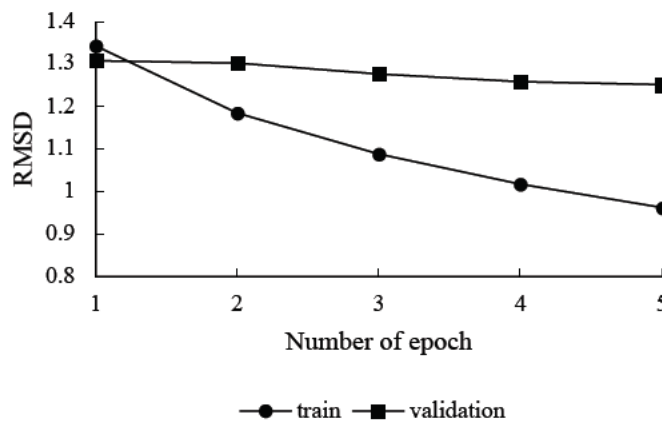


Figure 6.12: The learning curve of ANN with a 0.7:0.15:0.15 ratio split for training, validation, and test datasets in construction

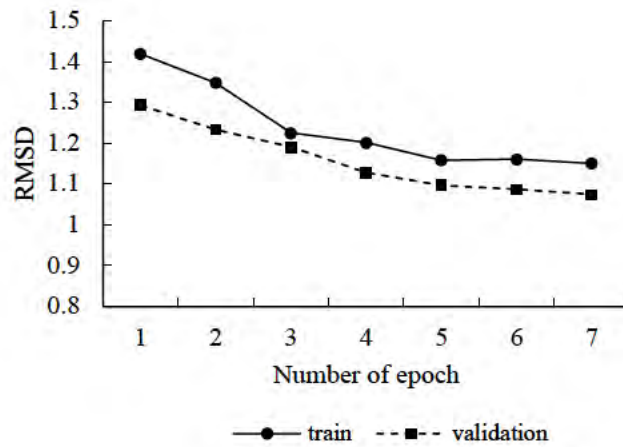


Figure 6.13: The learning curve of ANN with a 0.5:0.25:0.25 ratio split for training, validation, and test datasets in construction

The optimized parameters of RF were listed in Table 6.10.

Table 6.10: The optimized parameters of RF in construction

No.	parameters	values of optimized parameters	
		0.7:0.15:0.15	0.5:0.25:0.25
1	n estimators	235	231
2	min samples split	3	4
3	min samples leaf	3	3
4	bootstrap	False	False

The optimized parameters of SVC were listed in Table 6.11.

Table 6.11: The optimized parameters of SVC in construction

No.	parameters	values of optimized parameters	
		0.7:0.15:0.15	0.5:0.25:0.25
1	C	5.07	5.45
2	degree	7.00	7.00
3	gamma	0.03	0.05
4	kernel	sigmoid	sigmoid

The optimized parameters of ANN were listed in Table 6.12.

Table 6.12: The optimized parameters of ANN in construction

No.	parameters	values of optimized parameters	
		0.7:0.15:0.15	0.5:0.25:0.25
1	number of layers	3	3
2	number of neurons	18	19
3	activation function	relu	relu
4	dropout layers	1	1

6.4.3 Comparison of Based Model Performance

Based on the analysis of confusion matrices and result reports, including accuracy, recall, and F1 scores derived from the test sets of the three models, certain consistent conclusions can be drawn despite the utilization of different datasets in the construction set. Firstly, RF and SVC performed better when trained with an enormous amount of data or when dealing with imbalanced data among the four classes. In contrast, the performance of ANN depended more on the distribution of data within each category. Secondly, the ANN model trained with 50% of the data exhibited the best performance, achieving an overall accuracy of 0.77, surpassing RF (0.53) and SVC (0.52). Thirdly, the highest-performing class was still the fourth risk level, with an F1 score of 0.85, followed by class 1 with an F1 score of 0.83. Fourthly, classes 2 and 3, which had relatively closer risk scores, continued to show lower classification performance, with F1 scores of 0.72 and 0.61, respectively.

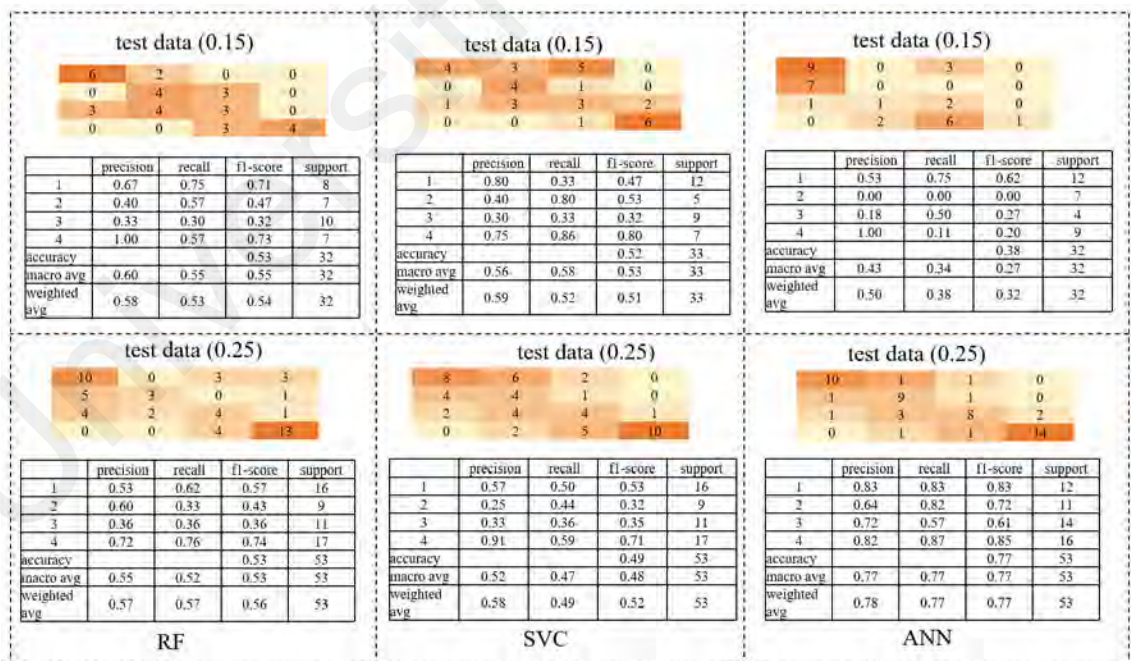


Figure 6.14: Confusion matrix and result reports of three models with two ratios split for training, validation, and test datasets in construction

In the ROC curves of the three models (Figure 6.15), it can be observed that for classes 4 and 1, except for the ANN model trained with 70% of the data, they were all located

closest to the upper-left corner of the graph. It indicated that higher True Positive Rates (TPR) and lower False Positive Rates (FPR), i.e., higher sensitivity and lower misclassification rate, resulted in better classification performance. On the other hand, classes 2 and 3, particularly class 3, were closer to the diagonal line $y=x$. The closer the points were to this line, the worse the classification results, approaching the random guessing result of 0.5.

The AUC value of the ANN model trained with the data ratio 0.5:0.25:0.25 reached the highest among all the models, with a value of 0.83. It indicated that the model distinguished between positive and negative instances well. However, for the ANN model trained with the imbalanced class data ratio of 0.7:0.15:0.15, the AUC value was the lowest at 0.42. The poor performance can be attributed to the severe class imbalance, where the model struggled to classify the minority classes effectively. The AUC value being even smaller than the result of random guessing (0.5) further highlighted the model's inadequate ability to make meaningful predictions.

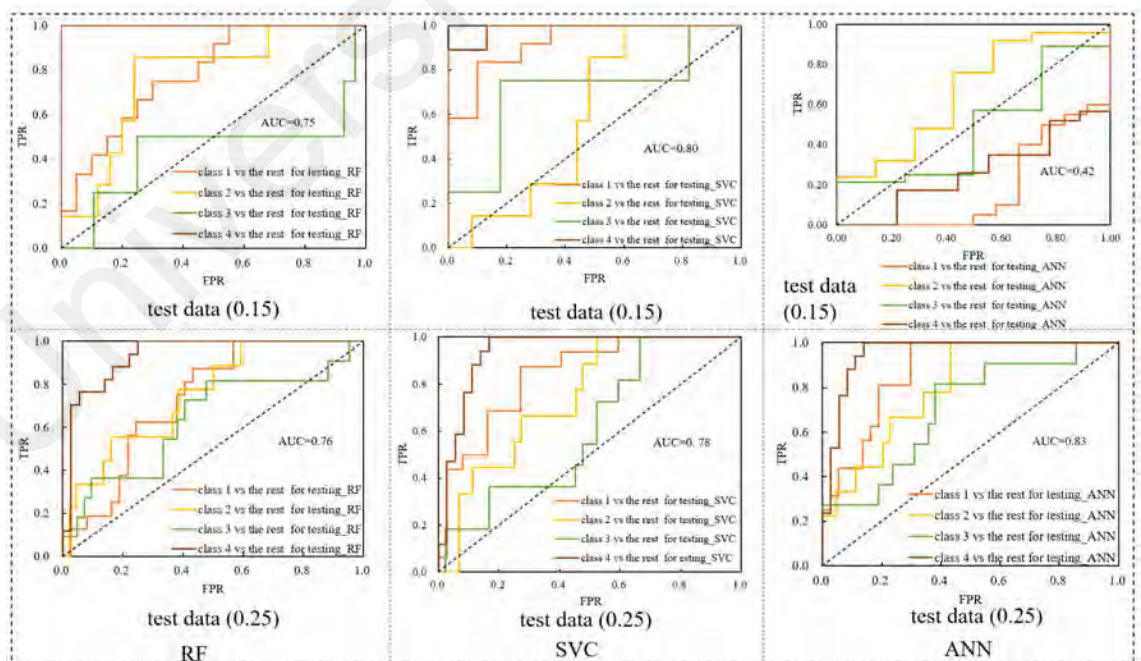


Figure 6.15: The ROC curves and AUC values of three models with two ratios split for training, validation, and test datasets in construction

6.4.4 Comparison Integrated Models Performance

Similar to the driving scenario experiments, both integration models demonstrated improved performance compared to the ANN model. For example, the weight computation method enhanced the overall model accuracy from 0.77 to 0.91, while the stacking method increased it to 0.88. Furthermore, for the challenging-to-classify classes 2 and 3, the weight computation method yielded higher F1 scores than ANN, with values of 0.83 and 0.88, respectively, as opposed to ANN's 0.72 and 0.61. Applying the stacking technique also increased F1 scores for classes 2 and 3, reaching 0.80 and 0.88. The weight computation method exhibited a slightly better improvement in model classification performance over the stacking technique, consistent with the results obtained in the first experiment. Additionally, for the more challenging classes 2 and 3, the classification results with high precision and low recall reappeared in the construction scene, as observed in the driving scenario.

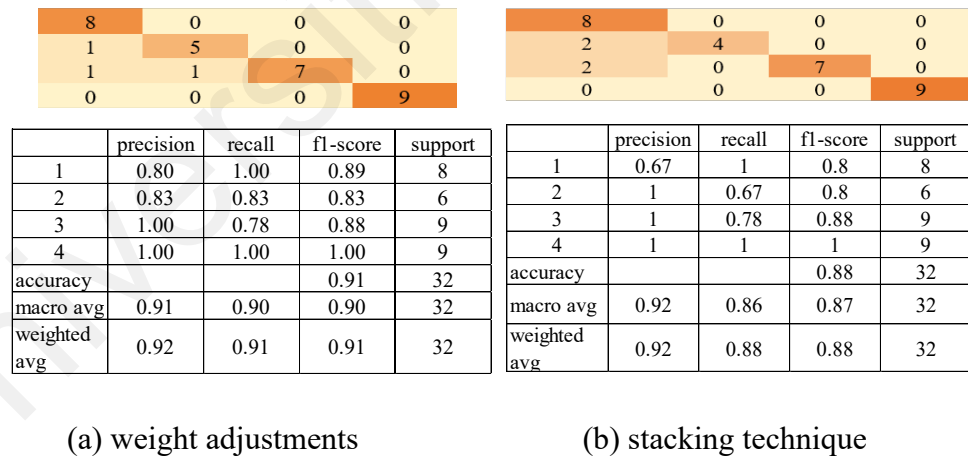


Figure 6.16: Confusion matrix and result reports of integrated model in constructing

Similarly, in the validation experiment, the AUC value under the stacking approach was 0.95, surpassing the AUC of ANN, which was 0.83. It also indicated that the integrated model outperformed the performance of an individual best-performing model.

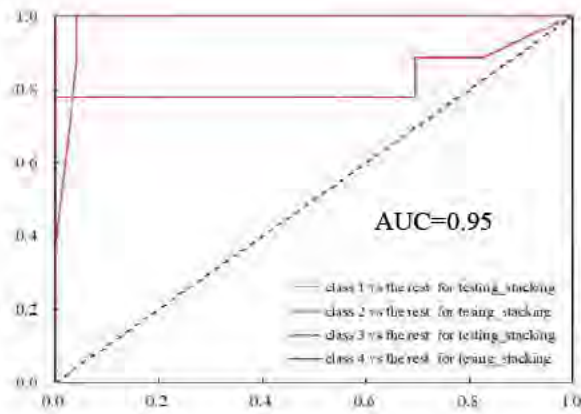


Figure 6.17: The ROC curves and AUC values of the integrated model with stacking technique in construction

6.5 Comparison and Discussion with Previous Machine Learning

The confirmation of the perceived risk classification for the construction scenario was found to be identical to that of the driving scenario. The following similarities were observed by comparing the results of model performance, indicator selection, and feature importance calculations with the previous analysis.

Firstly, regarding feature importance, the two experiments' results were different. Consistently, emotional response indicators such as A and D, instantaneous heart rate variability indicators like A++ and B--, and skin conductance-related indicator EDR were consistently identified as more critical classification features. However, in the construction stimulus, the increased complexity of risk categorization led to the inclusion of additional features, namely VLF, LF, LF/HF, and SD2. Nonetheless, their feature importance rankings were consistently lower than the abovementioned indicators. It emphasized once again that as risk levels became closer in magnitude, the challenge of risk classification naturally intensified, necessitating the incorporation of more features to enhance classification performance. However, these subsequently added features exhibited relatively lower importance rankings.

Secondly, both scenarios demonstrated that the ANN model outperformed RF and SVC when using balanced training data. However, the weighted adjustment integrated algorithm, based on the three trained models, surpassed ANN. This superior performance

was evident not only in classifying high and low-risk instances but also in situations involving closer risk degrees, such as situations 2 and 3. Furthermore, the ANN classifier outperformed a comparable study in high and low-risk classification (B. G. Lee et al., 2021). The F1 scores for low (0.83) and high-risk (0.85) exceeded those reported in the related research (0.819 for low-risk and 0.805 for high-risk). Naturally, the two integrated algorithms were also better than the previously trained model.

Thirdly, the two experiments yielded slightly inconsistent results when using RFE to select the optimal feature combinations. The former scenario identified 8 features, while the latter identified 10 features. This consistent observation was that the remaining features mostly showed significant differences among the four risk levels. However, the discrepancy could be attributed to the fact that in the latter scenario, risk scores for situations 2 and 3 were closer, making the classification more challenging and necessitating the retention of more features. Moreover, when comparing the selected features between the two scenarios, it was observed that ULF did not appear in the construction scenario. The correlation matrix heatmap indicated a strong positive correlation between ULF and VLF. While retaining VLF, the highly correlated feature ULF was discarded. P did not appear in the validation scenario, possibly due to its reduced significance in a scenario with a slightly higher average risk score. In addition, in the second scenario, additional HRV-related indicators, namely HR, LF/HF, LF, and SD2, were included. It suggested the vital value of heart rate-related indicators in more complex and challenging scenarios. This observation was further supported by the comparative analysis of the selected features in the two experiments, where besides emotional response and skin conductance-related features, all other features were HRV-related indicators.

6.6 Summary

The paradigm of assessing risk perception has been extracted and applied in a different scenario. The main finds and sights were summarized in Table 6.13.

Table 6.13: The main outcomes and limitations obtained from comparing the first and validated experimental analysis results

No.	Main outcomes and insights	Limitations
1	When the experimental scenario was changed from driving to construction, a comparative analysis of the data collected from these two distinct environments provided preliminary insights into the feasibility of using physiological and emotional response data to grade perceived risk.	1) Although both experiments were designed with scenarios of similar risk levels, the precise threshold at which risk levels are considered similar remains unresolved. 2) The risk perception assessment model proposed in this study heavily relies on changes in emotional and physiological indicators. If the risk scenarios are too similar, leading to subtle changes in these indicators, it significantly impacts the classification ability of machine learning models.
2	The comparison between statistical analysis and machine learning modeling showed some discrepancies. However, the results were consistent on key issues, such as trends in emotional and physiological changes, the regulation mechanisms of the autonomic nervous system, the sensitivity of different indicators, and the performance of various models. This consistency strongly supports the feasibility of using emotional and physiological changes as a basis for risk perception.	3) The model developed in this study can distinguish between different risk scenarios, but it only provides results in the form of discrete grades rather than continuous numerical values.

Statistical analysis confirmed the main consistent findings when comparing the construction and driving experiments, including similarities in emotional experiences, physiological changes and natural mediation. There were several dominating consensuses were emphasized again. As the risk level escalated from relative safety to higher degrees of risk, the elicited emotions transited from calm and tranquil to progressively negative. In response to risky situations, the body experiences an immediate increase in fundamental arousal levels, accompanied by a heightened sensation of powerlessness. Concurrent physiological responses further corroborated the elevation in arousal levels, evidenced by pronounced heart rate fluctuations between successive heartbeats. While the overarching pattern suggested enhanced SNS activity and a gradual decline in PNS activity, at a micro-level, both the SNS and PNS experienced augmentation, with the degree of SNS enhancement surpassing that of the PNS. It reflected the brain's continuous perception of external environmental changes and the ongoing adjustment of individual

risk perception during information assimilation. Simultaneously, the brain continually fine-tuned the heart's functioning to adapt to shifts in the external environment. Indeed, there were also some inconsistencies in the findings. For example, VLF did not exhibit significant differences in the construction scenario, and the discriminative ability of D within the construction setting was slightly diminished. These disparities primarily arose from variations in the magnitude of risk values across the four risk contexts in the two scenarios. Importantly, these differences did not affect the critical conclusions mentioned above.

Furthermore, features with significant differences or weak correlations with other features are more likely to remain. Additionally, the PAD emotion model can be employed to represent the emotions induced by risk situations quantitatively, and the roles of the three dimensions must be considered when evaluating subjective risk. As for physiological indicators, those capable of reflecting instantaneous changes play a more crucial role in classification. Considering time efficiency and computational costs, specific indicators, such as HRV in the time domain, skin temperature, and pupil diameter, can be omitted.

Finally, the weighted adjustment integrated algorithm outperforms ANN in evaluating inherent and subjective risk characteristics. ANN surpasses RF and SVC and performs well in high and low-risk situations. However, the weighted adjustment integrated algorithm excels in classifying risk situations with closer risk levels. Therefore, the comparative analysis of the two experiments demonstrates the excellent generalization ability of the weighted adjustment integrated technique, making it a more valuable tool for transfer and utilization in various scenarios.

CHAPTER 7: CONCLUSION AND RECOMMENDATION

7.1 Conclusions

The process of assessing individual perceived risk occurs within the human brain and is characterized by subjectivity, opacity, and inherent nature. The risk perception process is influenced by two fundamentally different modes: intuition and analysis, which exert varying proportions of impact on risk perception. Hence, this study adopts a paradigm that diverges from traditional engineering risk assessment methodology to quantify perceived risk. The research findings, confirmed through two similar experiments, can be summarized into three key points.

7.1.1 Characterization of physiological and emotional changes in risky situations

As the risk level increases, individuals experience intensified discomfort, heightened arousal, and a stronger feeling of losing control over the danger. These three dimensions-pleasure, arousal and dominance- can distinguish various risk situations. However, the effectiveness of their discriminative power depends on the extent of differences in risk degrees. All three dimensions demonstrate distinguishing capabilities in situations with a significant difference in risk levels, such as between safe and risky situations. Conversely, when moderate risk level differences exist, dominance (D) and arousal (A) exhibit more substantial discriminative power. In cases where the risk levels are relatively close, A and P demonstrate better discriminative ability than D.

Physiological responses during the risk perception process resemble emotional fluctuations. As the perceived risk intensifies, skin conductance response increases, heart rate decreases, and there are significant fluctuations in instantaneous heart rate. Although changes in eye activity and skin temperature also occur, they do not reach statistical significance. Similarly, physiological metrics exhibit more substantial value differences when risk disparities are significant. However, as the risk levels become closer, the magnitude of physiological changes decreases. Nevertheless, some indicators still reflect

short-term physiological changes, such as EDR, HR, LF/HF, A++, and B--, which consistently demonstrate exceptional discriminative power. These metrics can be considered sensitive indicators of risk perception.

The statistical differences in emotional or physiological responses in the present study are primarily attributed to varying risk levels, with individual differences accounting for a minor proportion. This phenomenon may be attributed to the fact that the participants recruited for the two experiments exhibit proximity regarding age, knowledge, experience, and skills.

The statistical analysis of physiological indicators like skin conductance, heart rate variability, pupil diameter, and skin temperature allow us to understand the underlying physiological mechanisms during a risk perception process. In a relatively safe situations, the sympathetic and parasympathetic nervous systems are in relative balance. As the risk level increases, this balance is disrupted, leading to enhanced sympathetic activity and reduced parasympathetic activity. At the highest level, both systems exhibit peak activity. These changes aim to motivate humans to escape or deal with risky situations. However, it is essential to note that the two antagonistic branches of the autonomic nervous system are simultaneously active, and the magnitude of changes in sympathetic and parasympathetic activity does not strictly follow a linear relationship with the growth of risk levels. According to the Free-Energy Principle, reciprocal information exchange occurs between the visceral organs and the brain at the micro-level. The brain continuously adjusts its risk perception assessment based on the information from the internal organs and the external environment. During this process, both sympathetic and parasympathetic activities increase and undergo rapid changes, but overall, the increase in sympathetic activity is more pronounced than that of the parasympathetic nervous system. These physiological changes reflect the complex and dynamic nature of the inner risk perception process, where the body's responses to risk are influenced by intricate

interactions of various physiological factors rather than adhering to superficial linear relationships.

7.1.2 Machine learning model to assess subjective perceived risk

Using ANN effectively classifies situations with substantial differences in risk levels, such as between high and low-risk. However, when risk degree differences are minimal, the nonlinear fitting capability of ANN falls short of the requirements. In such cases, there is no need for highly complex deep learning algorithms. Instead, an integration technique based on adjusting the weight coefficients of three base models can address this challenge. The fundamental concept is to increase the weights of base models that correctly classify while reducing the weights of those that classify incorrectly. This concept, while straightforward, offers a solution to a challenging problem.

Features with statistical differences or more considerable variations in a short-time are more likely to be retained. In particular, the roles of these three dimensions (P, A and D) deserve attention when classifying perceived risks. At the same time, the instantaneous physiological responses associated with risk perception often exhibit favorable feature importance.

7.1.3 Validate the performance of the proposed machine learning model

The experimental findings confirming construction scenarios are fundamentally congruent with driving experiments, even though the data analysis results exhibit some discrepancies in terms of statistical outcomes, specific performance metrics, chosen indicators, and feature importance. For instance, emotional and physiological response indicators demonstrated different discriminative abilities across various risk scenarios, yet they reaffirmed the presence of the exact neuromodulation mechanisms. Similarly, the dissimilarities for these indicators with significant differences are primarily attributed to variations in risk situations, with individual differences playing a minor role. Furthermore, the integrated model, adjusting class weights on the base model,

demonstrates superior classification performance in situations where risk degrees are closely aligned. Meanwhile, the ANN shows enough ability to distinguish risk situations with significant disparities. Both of these conclusions were confirmed during the validation experiments. Additionally, physiological indicators with no significant differences and minor variations are more likely to be removed during the feature selection. Additionally, the consistent observation of the relatively high importance of emotional and instantaneous physiological response features underscores the agreement among these findings.

7.2 Significance of Study

7.2.1 Significance of Knowledge

The present research provides a new perspective to understand risk perception deeply through the application of physiological and emotional measurement techniques. The simultaneous physiological and emotional reactions served as a direct and scientific approach to reflect the internal process associated with perceiving risk. The underlying central nervous mediation mechanisms are revealed by capturing and analysing emotional and physiological data across diverse risk conditions. These findings contribute to the advancement of our comprehension of risk psychology and neuroscience. More importantly, this study further refines the paradigm of assessing perceived risk through machine learning algorithms using physiological and emotional responses.

7.2.2 Significance of Application

In the practical domain, the present study demonstrates that the machine learning model perform well in assessing perceived risk from the individual aspect. It provides a tool that allows workers to ensure that their true feelings about risk are taken into account. They will feel secure when they have strong confidence that the risk is under control. This valuable trained model can provide individuals in risky situations with objective and authentic insights into risk perception, enhancing safety management professionals'

understanding of risk scenarios beyond traditional engineering risk assessments. On one hand, this model can assist scientifically in identifying individuals with incorrect risk perception capabilities for high-risk positions. On the other hand, it assists safety managers in recognizing potential high-risk places within the workplace.

Workers' risk perception can enhance safety managers' analyses. Regardless of whether the risk is overestimated or underestimated, safety professionals need to reevaluate risk, considering the combination of engineering risk and perceived risk. By synthesizing these two aspects, it becomes possible to discern whether the issue lies with individuals exhibiting risk perception biases or if there are genuine inaccuracies in the engineering risk assessment. The former case necessitates strengthening attitudes, knowledge, and skills through safety training interventions. In the latter, it calls for improvements in existing risk mitigation techniques to reduce risks to an acceptable level. In brief, the assessment of perceived risk, combined with engineering risk assessment results, facilitates the development of more effective risk management strategies and measures. It also paves the way for transitioning from Safety I to Safety II, with a greater emphasis on human factors within complex and dynamically changing contexts.

7.3 Limitations

The physiological and emotional responses variations across different risk-level contexts form the foundation of machine learning model performance. Nevertheless, individuals working frequently in high-risk positions may have become accustomed to high-risk work environments, potentially leading to minimal fluctuations in their physiological signals. This limitation constrains the classification capability of machine learning models.

7.4 Recommendations for Future Work

(a) Recruiting Diverse Participants

The proportion of individual differences in specific indicators of physiological and emotional measurements contributing to the overall variance is relatively small, potentially attributed to controlled subject selection during recruitment. Subsequent research endeavors may consider relaxing subject selection criteria by including participants with varying characteristics, such as personality, age, knowledge, and skills. This approach would facilitate investigations into the origins of differences in the studied indicators and the development of classification models.

(b) Using virtual reality (VR) as a stimulus

Virtual reality scenarios can be considered risk situations, as they offer a closer approximation to real-world contexts. This stimulus can lead to participants' physiological and emotional responses more aligned with actual experiences.

(c) Establishment of a universal database

It is necessary to establish a universal database of physiological and emotional responses to develop assessment models that are both comparative and perform optimally using the same dataset. The data used in subjective risk assessments mostly come from researchers' experiments. Due to variations in specific risk situations and levels, the performance of assessment models needs to be more comparable across different scenarios, limiting the generalizability of the models.

(d) Adding EEG measurement

Nonlinear HRV indices have demonstrated pronounced fluctuations at the micro-level. It is necessary to add an EEG measurement. It is primarily due to the human brain's integral role in the autonomic nervous system and its superior temporal resolution, enabling further analysis of the underlying physiological mechanisms involved in risk perception at a microscopic or fine-grained level. Indeed, a high temporal and spatial

resolution of a signal generates a larger volume of data, posing challenges in selecting subsequent machine learning models.

Universiti Malaya

REFERENCES

- Adem, A., Cakit, E., & Dagdeviren, M. (2020). Occupational health and safety risk assessment in the domain of Industry 4.0. *Sn Applied Sciences*, 2(5).
- Agyeiwaah, E., Adam, I., Dayour, F., & Badu Baiden, F. (2021). Perceived impacts of COVID-19 on risk perceptions, emotions, and travel intentions: evidence from Macau higher educational institutions. *Tourism Recreation Research*, 46(2), 195-211.
- Ahmed, T., Qassem, M., & Kyriacou, P. A. (2022). Physiological monitoring of stress and major depression: A review of the current monitoring techniques and considerations for the future. *Biomedical signal processing and control*, 75.
- Aiello, G., Catania, P., Vallone, M., & Venticinque, M. (2022). Worker safety in agriculture 4.0: A new approach for mapping operator's vibration risk through Machine Learning activity recognition. *Computers and Electronics in Agriculture*, 193, 106637.
- Alrawad, M., Lutfi, A., Alyatama, S., Elshaer, I. A., & Almaiah, M. A. (2022). Perception of occupational and environmental risks and hazards among mineworkers: A psychometric paradigm approach. *International Journal of Environmental Research and Public Health*, 19(6), 3371.
- Alrefaie, M. T., Summerskill, S., & Jackson, T. W. (2019). In a heart beat: Using driver's physiological changes to determine the quality of a takeover in highly automated vehicles. *Accident Analysis and Prevention*, 131, 180-190.
- Antwi-Afari, M. F., Anwer, S., Umer, W., Mi, H.-Y., Yu, Y., Moon, S., & Hossain, M. U. (2023). Machine learning-based identification and classification of physical fatigue levels: A novel method based on a wearable insole device. *International Journal of Industrial Ergonomics*, 93, 103404.
- Arezes, P. M., & Miguel, A. S. (2008). Risk perception and safety behaviour: A study in an occupational environment. *Safety Science*, 46(6), 900-907.

- Armougum, A., Gaston-Bellegarde, A., Joie-La Marle, C., & Piolino, P. (2020). Physiological investigation of cognitive load in real-life train travelers during information processing. *Applied ergonomics*, 89, 103180.
- Armour, J. A. (2003). Neurocardiology. *Anatomical and Functional Principles*. Boulder Creek, CA: HeartMath Research Center, Institute of HeartMath, Publication(03-011).
- Aryal, A., Ghahramani, A., & Becerik-Gerber, B. (2017). Monitoring fatigue in construction workers using physiological measurements. *Automation in Construction*, 82, 154-165.
- Atik, O., & Arslan, O. (2019). Use of eye tracking for assessment of electronic navigation competency in maritime training. *Journal of Eye Movement Research*, 12(3).
- Aven, T. (2016). Risk assessment and risk management: Review of recent advances on their foundation. *European Journal of Operational Research*, 253(1), 1-13.
- Aven, T. (2018). How the integration of System 1-System 2 thinking and recent risk perspectives can improve risk assessment and management. *Reliability Engineering & System Safety*, 180, 237-244.
- Aven, T., & Kristensen, V. (2005). Perspectives on risk: review and discussion of the basis for establishing a unified and holistic approach. *Reliability Engineering & System Safety*, 90(1), 1-14.
- Aven, T., & Renn, O. (2010). *Risk management and governance: Concepts, guidelines and applications* (Vol. 16). Springer Science & Business Media.
- Aven., & Terje. (2022). A risk science perspective on the discussion concerning Safety I, Safety II and Safety III. *Reliability Engineering & System Safety*, 217, 108077.

- Awais, M., Badruddin, N., & Drieberg, M. (2017). A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and wearability. *Sensors*, *17*(9), 1991.
- Ayodele, T. O. (2010). Types of machine learning algorithms in new advances in machine learning. In. croatia, rijeka: InTech.
- Azadeh, A., & Zarrin, M. (2016). An intelligent framework for productivity assessment and analysis of human resource from resilience engineering, motivational factors, HSE and ergonomics perspectives. *Safety Science*, *89*, 55-71.
- Bačić, D., & Henry, R. (2022). Advancing our understanding and assessment of cognitive effort in the cognitive fit theory and data visualization context: Eye tracking-based approach. *Decision support systems*, *163*, 113862.
- Bagley, L., Boag-Hodgson, C., & Stainer, M. (2023). Hero or hazard: A systematic review of individual differences linked with reduced accident involvement and influencing success during emergencies. *Heliyon*, *9*(4), e15006.
- Baldauf, D., Burgard, E., & Wittmann, M. (2009). Time perception as a workload measure in simulated car driving. *Applied ergonomics*, *40*(5), 929-935.
- Barnard, M. P., & Chapman, P. (2016). Are anxiety and fear separable emotions in driving? A laboratory study of behavioural and physiological responses to different driving environments. *Accident Analysis and Prevention*, *86*, 99-107.
- Bell, J., & Holroyd, J. (2009). Review of human reliability assessment methods, RR679. *Health and Safety Executive*.
- Bellamy, L. J., Geyer, T. A. W., & Wilkinson, J. (2008). Development of a functional model which integrates human factors, safety management systems and wider

organisational issues. *Safety Science*, 46(3), 461-492.

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190(1), 80-91.

Bernhardt, K. A., Poltavski, D., Petros, T., Ferraro, F. R., Jorgenson, T., Carlson, C., Drechsel, P., & Iseminger, C. (2019). The effects of dynamic workload and experience on commercially available EEG cognitive state metrics in a high-fidelity air traffic control environment. *Applied ergonomics*, 77, 83-91.

Berntson, G. G., & Cacioppo, J. T. (1999). Heart rate variability: a neuroscientific perspective for further studies. *Cardiac Electrophysiology Review*, 3, 279-282.

Berntson, G. G., Thomas Bigger Jr, J., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., Nagaraja, H. N., Porges, S. W., Saul, J. P., & Stone, P. H. (1997). Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology*, 34(6), 623-648.

Bevilacqua, M., & Ciarapica, F. E. (2018). Human factor risk management in the process industry: A case study. *Reliability Engineering & System Safety*, 169, 149-159.

Bevilacqua, M., Ciarapica, F. E., & Giacchetta, G. (2010). Data mining for occupational injury risk: a case study. *International Journal of Reliability, Quality and Safety Engineering*, 17(04), 351-380.

Bhat, G., Deb, R., Chaurasia, V. V., Shill, H., & Ogras, U. Y. (2018). Online human activity recognition using low-power wearable devices. 2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD),

Bhatia, S. (2019). Predicting risk perception: New insights from data science. *Management Science*, 65(8), 3800-3823.

- Billman, G. E. (2013). The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. In (Vol. 4, pp. 26): Frontiers Media SA.
- Billman, G. E., Huikuri, H. V., Sacha, J., & Trimmel, K. (2015). An introduction to heart rate variability: methodological considerations and clinical applications. In (Vol. 6, pp. 55): Frontiers Media SA.
- Biondi, F. N., Saberi, B., Graf, F., Cort, J., Pillai, P., & Balasingam, B. (2023). Distracted worker: Using pupil size and blink rate to detect cognitive load during manufacturing tasks. *Applied ergonomics*, *106*, 103867.
- Bjørneseth, F. B., Clarke, L., Dunlop, M., & Komandur, S. (2014). Towards an understanding of operator focus using eye-tracking in safety-critical maritime settings. International Conference on Human Factors in Ship Design & Operation,
- Bly, M. (2011). *Deepwater Horizon accident investigation report*. Diane Publishing.
- Botterill, L., & Mazur, N. (2004). Risk and risk perception: A literature review. *Kingstrom, ACT: Australian Government Rural Industries Research and Development Corporation*.
- Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Society for psychophysiological research ad hoc committee on electrodermal measures. Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017-1034.
- Bouyer, M., Bagdassarian, S., Chaabanne, S., & Mullet, E. (2001). Personality correlates of risk perception. *Risk Analysis*, *21*(3), 457-466.
- Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602-607.
- Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology*, *49*(1), 1017-1034.

Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.

Brennan, M., Palaniswami, M., & Kamen, P. (2001). Do existing measures of Poincare plot geometry reflect nonlinear features of heart rate variability? *Ieee Transactions on Biomedical Engineering*, 48(11), 1342-1347.

Bronfman, N. C., Cifuentes, L. A., & Gutiérrez, V. V. (2008). Participant - focused analysis: Explanatory power of the classic psychometric paradigm in risk perception. *Journal of Risk Research*, 11(6), 735-753.

Brookhuis, K. A., & de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident Analysis and Prevention*, 42(3), 898-903.

Brown, V. J. (2014). Risk Perception: It's Personal. *Environmental Health Perspectives*, 122(10), A276-A279.

Brun, W. (1994). Risk perception: Main issues, approaches and findings. *Subjective probability*, 295-320.

Budidha, K., & Kyriacou, P. A. (2019). Photoplethysmography for Quantitative Assessment of Sympathetic Nerve Activity (SNA) During Cold Stress. *Frontiers in Physiology*, 9.

Burns, W. J., & Slovic, P. (2012). Risk perception and behaviors: anticipating and responding to crises. *Risk Analysis*.

Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Vogt-Maranto, L., & Zdeborová, L. (2019). Machine learning and the physical sciences. *Reviews of Modern Physics*, 91(4), 045002.

- Carriço, A., Gomes, A. R., & Gonçalves, A. P. (2015). Quantitative analysis of the construction industry workers' perception of risk in municipalities surrounding Salvador. *Procedia Manufacturing*, 3, 1846-1853.
- Čegovnik, T., Stojmenova, K., Jakus, G., & Sodnik, J. (2018). An analysis of the suitability of a low-cost eye tracker for assessing the cognitive load of drivers. *Applied ergonomics*, 68, 1-11.
- Cha, Y.-J. (2000). Risk perception in Korea: An application of psychometric paradigm. *International Journal of Risk Assessment and Management*, 1(1-2), 42-51.
- Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. Guilford Press.
- Chan, J. Y.-L., Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z.-W., & Chen, Y.-L. (2022). Mitigating the multicollinearity problem and its machine learning approach: a review. *Mathematics*, 10(8), 1283.
- Chang, Y. H., Yang, H. H., & Hsiao, Y. J. (2016). Human risk factors associated with pilots in runway excursions. *Accident Analysis and Prevention*, 94, 227-237.
- Charlton, S. G., Starkey, N. J., Perrone, J. A., & Isler, R. B. (2014). What's the risk? A comparison of actual and perceived driving risk. *Transportation research part F: traffic psychology and behaviour*, 25, 50-64.
- Chauvin, B., Hermand, D., & Mullet, E. (2007). Risk perception and personality facets. *Risk Analysis: An International Journal*, 27(1), 171-185.
- Choi, B., Jebelli, H., & Lee, S. (2019). Feasibility analysis of electrodermal activity (EDA) acquired from wearable sensors to assess construction workers' perceived risk. *Safety Science*, 115, 110-120.

- Choi, B., Jebelli, H., & Lee, S. (2019). Feasibility analysis of electrodermal activity (EDA) acquired from wearable sensors to assess construction workers' perceived risk. *Safety science*, *115*, 110-120.
- Choi, B., Lee, G., Jebelli, H., & Lee, S. (2019). Assessing workers perceived risk during construction task using a wristband-type biosensor. *arXiv preprint arXiv:1908.05133*.
- Choi, B., & Lee, S. (2018). An Empirically Based Agent-Based Model of the Sociocognitive Process of Construction Workers' Safety Behavior. *Journal of Construction Engineering and Management*, *144*(2).
- Choi, M., Park, M., Lee, H.-S., Hwang, S., Anderson, K., & Lee, S. (2016). Understanding the role of dynamic risk perception during fire evacuations using agent-based modeling. Construction Research Congress 2016,
- Choi, M. K., & Seong, P. H. (2020). A methodology for evaluating human operator's fitness for duty in nuclear power plants. *Nuclear Engineering and Technology*, *52*(5), 984-994.
- Christian, M. S., Bradley, J. C., Wallace, J. C., & Burke, M. J. (2009). Workplace Safety: A Meta-Analysis of the Roles of Person and Situation Factors. *Journal of Applied Psychology*, *94*(5), 1103-1127.
- Christopher, A. A., & alias Balamurugan, S. A. (2014). Prediction of warning level in aircraft accidents using data mining techniques. *The Aeronautical Journal*, *118*(1206), 935-952.
- Coeugnet, S., Naveteur, J., Antoine, P., & Anceaux, F. (2013). Time pressure and driving: Work, emotions and risks. *Transportation Research Part F-Traffic Psychology and Behaviour*, *20*, 39-51.

- Committee, R. R. (1980). Accidents in Norway. How Do We Perceive and Handle Risk. *Royal Norwegian Council for Scientific and Industrial Research (NTNF), Oslo.*
- Connelly, C. (1997). Changes in operator activities due to upsets. *Newsletter, Dayton, OH: Beville Engineering.*
- Cooke, L. (2005). Eye tracking: How it works and how it relates to usability. *Technical Communication, 52(4), 456-463.*
- Cristea, M., & Delhomme, P. (2016). The effects of co-presence on risk perception and intention to engage in risky behaviors. *Journal of Safety Research, 56, 97-103.*
- Cui, F., Liu, Y., Chang, Y., Duan, J., & Li, J. (2016). An overview of tourism risk perception. *Natural Hazards, 82, 643-658.*
- Curiel-Ramirez, L. A., Ramirez-Mendoza, R. A., Carrera, G., Izquierdo-Reyes, J., & Bustamante-Bello, M. R. (2019). Towards of a modular framework for semi-autonomous driving assistance systems. *International Journal of Interactive Design and Manufacturing - Ijidem, 13(1), 111-120.*
- Curtis, P., Carey, M., & Committee of Sponsoring Organizations of the Treadway Commission. (2012). *Risk assessment in practice.*
- Damasio, A. (1994). Descartes' error: Emotion, rationality and the human brain. *New York: Putnam, 352.*
- DASDEMİR, Y., Yildirim, E., & Yildirim, S. (2017). Emotion analysis using different stimuli with EEG signals in emotional space. *Natural and Engineering Sciences, 2(2), 1-10.*
- Dawson, M. E., Schell A.,M., & Filion, D.L. (2007). The electrodermal system. In L. G. T. In J.T.Cacioppo, & G.G. Berntson (Ed.), *Handbook of psychophysiology* (3 ed.,

pp. 159-181). Cambridge University Press.

Dawson, M. E., Schell, A. M., & Courtney, C. G. (2011). The skin conductance response, anticipation, and decision-making. *Journal of Neuroscience, Psychology, and Economics*, 4(2), 111.

Dawson, M. E., Schell, A. M., Filion, D. L., Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (2007). *Handbook of psychophysiology*. NY: Cambridge University Press.

De Salvo, M., Grilli, G., Notaro, S., & Signorello, G. (2022). Do risk perception and safety of sites influence rock climbing destination choices? *Journal of Outdoor Recreation and Tourism*, 37, 100486.

Deacon, T., Amyotte, P. R., & Khan, F. I. (2010). Human error risk analysis in offshore emergencies. *Safety Science*, 48(6), 803-818.

Deery, H. A. (1999). Hazard and risk perception among young novice drivers. *Journal of Safety Research*, 30(4), 225-236.

Deffenbacher, J. L., Deffenbacher, D. M., Lynch, R. S., & Richards, T. L. (2003). Anger, aggression, and risky behavior: a comparison of high and low anger drivers. *Behaviour Research and Therapy*, 41(6), 701-718.

Dehais, F., Causse, M., Vachon, F., & Tremblay, S. (2012). Cognitive conflict in human-automation interactions: a psychophysiological study. *Applied ergonomics*, 43(3), 588-595.

Demaree, H. A., DeDonno, M. A., Burns, K. J., Feldman, P., & Everhart, D. E. (2009). Trait dominance predicts risk-taking. *Personality and Individual Differences*, 47(5), 419-422.

Dettmers, C., Fatepour, D., Faust, H., & Jerusalem, F. (1993). Sympathetic Skin-

Response Abnormalities in Amyotrophic-Lateral-Sclerosis. *Muscle & Nerve*, 16(9), 930-934.

Devlin, S. P., Brown, N. L., Drollinger, S., Sibley, C., Alami, J., & Riggs, S. L. (2022). Scan-based eye tracking measures are predictive of workload transition performance. *Applied ergonomics*, 105, 103829.

Di Simplicio, M., Costoloni, G., Western, D., Hanson, B., Taggart, P., & Harmer, C. (2012). Decreased heart rate variability during emotion regulation in subjects at risk for psychopathology. *Psychological medicine*, 42(8), 1775-1783.

Dias, D., & Cunha, J. P. S. (2018). Wearable Health Devices-Vital Sign Monitoring, Systems and Technologies. *Sensors*, 18(8).

Digiesi, S., Manghisi, V. M., Facchini, F., Klose, E. M., Foglia, M. M., & Mummolo, C. (2020). Heart rate variability based assessment of cognitive workload in smart operators. *Management and Production Engineering Review*, 11(3), 56-64.

DiMattia, D. G., Khan, F. I., & Amyotte, P. R. (2005). Determination of human error probabilities for offshore platform musters. *Journal of Loss Prevention in the Process Industries*, 18(4-6), 488-501.

Ding, C., Wu, X. K., Yu, G. Z., & Wang, Y. P. (2016). A gradient boosting logit model to investigate driver's stop-or-run behavior at signalized intersections using high-resolution traffic data. *Transportation Research Part C-Emerging Technologies*, 72, 225-238.

Ding, H., Ghazilla, R. A. R., Singh, R. S. K., & Wei, L. (2022). Deep learning method for risk identification under multiple physiological signals and PAD model. *Microprocessors and Microsystems*, 88, 104393.

Ding, H., Raja Ghazilla, R. A., Kuldip Singh, R. S., & Wei, L. (2022). Vehicle driving risk prediction model by reverse artificial intelligence neural network. *Computational intelligence and neuroscience*, 2022(1), 3100509.

- Distefano, N., Leonardi, S., Pulvirenti, G., Romano, R., Boer, E., & Wooldridge, E. (2022). Mining of the association rules between driver electrodermal activity and speed variation in different road intersections. *IATSS research*, 46(2), 200-213.
- Doorley, R., Pakrashi, V., Byrne, E., Comerford, S., Ghosh, B., & Groeger, J. A. (2015). Analysis of heart rate variability amongst cyclists under perceived variations of risk exposure. *Transportation research part F: traffic psychology and behaviour*, 28, 40-54.
- Dovgalecs, V., Megret, R., Wannous, H., & Berthoumieu, Y. (2010). Semi-supervised learning for location recognition from wearable video. 2010 International Workshop on Content Based Multimedia Indexing (CBMI),
- Drač, S., & Ric, F. (2012). The effect of emotions on risk perception: Experimental evaluation of the affective tendencies framework. *Psihologija*, 45(4), 409-416.
- Dryhurst, S., Schneider, C. R., Kerr, J., Freeman, A. L., Recchia, G., Van Der Bles, A. M., Spiegelhalter, D., & Van Der Linden, S. (2022). Risk perceptions of COVID-19 around the world. In *COVID-19* (pp. 162-174). Routledge.
- Dyer, J., & Kolic, B. (2020). Public risk perception and emotion on Twitter during the Covid-19 pandemic. *Applied Network Science*, 5(1), 99.
- Earle, T. C., & Lindell, M. K. (1984). Public perception of industrial risks: a free-response approach. *Low-Probability High-Consequence Risk Analysis: Issues, Methods, and Case Studies*, 531-550.
- Embrey, D., Humphreys, P., Rosa, E., Kirwan, B., & Rea, K. (1984). *SLIM-MAUD: an approach to assessing human error probabilities using structured expert judgment. Volume I. Overview of SLIM-MAUD.*

- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human factors*, 37(1), 32-64.
- Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American psychologist*, 49(8), 709.
- Evin, M., Hidalgo-Munoz, A., Béquet, A. J., Moreau, F., Tattegrain, H., Berthelon, C., Fort, A., & Jallais, C. (2022). Personality trait prediction by machine learning using physiological data and driving behavior. *Machine Learning with Applications*, 9, 100353.
- Fan, X., Wang, F., Lu, Y., Song, D., & Liu, J. (2018). Eye gazing enabled driving behavior monitoring and prediction. 2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW),
- Fan, X., Zhao, C., Zhang, X., Luo, H., & Zhang, W. (2020). Assessment of mental workload based on multi-physiological signals. *Technology and Health Care*, 28(S1), 67-80.
- Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, 13(1), 1-17.
- Fischhoff, B., Slovic, P., Lichtenstein, S., Read, S., & Combs, B. (1978). How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sciences*, 9, 127-152.
- Fista, B., Azis, H., Aprilya, T., Saidatul, S., Sinaga, M., Pratama, J., Syalfinaf, F., & Amalia, S. (2019). Review of Cognitive Ergonomic Measurement Tools. IOP Conference Series: Materials Science and Engineering,
- Flin, R., & Mearns, K. (1994). Risk perception and safety in the offshore oil industry. SPE International Conference and Exhibition on Health, Safety, Environment, and Sustainability,

- Flin, R. H., O'Connor, P., & Crichton, M. (2008). *Safety at the sharp end: a guide to non-technical skills*. Ashgate Publishing, Ltd.
- Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in cognitive sciences*, 13(7), 293-301.
- Fu, M. Q., Liu, R., & Zhang, Y. (2021). Why do people make risky decisions during a fire evacuation? Study on the effect of smoke level, individual risk preference, and neighbor behavior. *Safety Science*, 140.
- Fung, I. W. H., Tam, V. W. Y., Lo, T. Y., & Lu, L. L. H. (2010). Developing a Risk Assessment Model for construction safety. *International Journal of Project Management*, 28(6), 593-600.
- Fyhri, A., & Phillips, R. O. (2013). Emotional reactions to cycle helmet use. *Accident Analysis and Prevention*, 50, 59-63.
- Gajardo, A. I. J., Madariaga, S., & Maldonado, P. E. (2019). Autonomic nervous system assessment by pupillary response as a potential biomarker for cardiovascular risk: A pilot study. *Journal of Clinical Neuroscience*, 59, 41-46.
- Gao, S., & Wang, L. (2020). Effects of mental workload and risk perception on pilots' safety performance in adverse weather contexts. *Engineering Psychology and Cognitive Ergonomics*. Cognition and Design: 17th International Conference, EPCE 2020,
- Garbey, M., Sun, N., Merla, A., & Pavlidis, I. (2007). Contact-free measurement of cardiac pulse based on the analysis of thermal imagery. *Ieee Transactions on Biomedical Engineering*, 54(8), 1418-1426.
- Gardner, G. T., & Gould, L. C. (1989). Public perceptions of the risks and benefits of

technology. *Risk Analysis*, 9(2), 225-242.

Gertman, D., Blackman, H., Marble, J., Byers, J., & Smith, C. (2005). The SPAR-H human reliability analysis method. *US Nuclear Regulatory Commission*, 230(4), 35.

Giagloglou, E., Mijovic, P., Brankovic, S., Antoniou, P., & Macuzic, I. (2019). Cognitive status and repetitive working tasks of low risk. *Safety Science*, 119, 292-299.

Gierlach, E., Belsher, B. E., & Beutler, L. E. (2010). Cross - cultural differences in risk perceptions of disasters. *Risk Analysis: An International Journal*, 30(10), 1539-1549.

Goh, Y. M., Ubeynarayana, C. U., Wong, K. L. X., & Guo, B. H. W. (2018). Factors influencing unsafe behaviors: A supervised learning approach. *Accident Analysis and Prevention*, 118, 77-85.

Golmohammadi, R., Darvishi, E., Motlagh, M. S., Faradmal, J., Aliabadi, M., & Rodrigues, M. A. (2022). Prediction of occupational exposure limits for noise-induced non-auditory effects. *Applied ergonomics*, 99, 103641.

Gonzalez, C., Thomas, R. P., & Vanyukov, P. (2005). The relationships between cognitive

Grassmann, M., Vlemincx, E., von Leupoldt, A., & Van den Bergh, O. (2017). Individual differences in cardiorespiratory measures of mental workload: An investigation of negative affectivity and cognitive avoidant coping in pilot candidates. *Applied ergonomics*, 59, 274-282.

Greco, A., Valenza, G., Lanata, A., Scilingo, E. P., & Citi, L. (2016). cvxEDA: A Convex Optimization Approach to Electrodermal Activity Processing. *Ieee Transactions on Biomedical Engineering*, 63(4), 797-804.

- Guan, X. Y., & Chen, C. (2014). Using social media data to understand and assess disasters. *Natural Hazards*, 74(2), 837-850.
- Gui, X. N., Kou, Y. B., Pine, K. H., & Chen, Y. A. (2017). Managing Uncertainty: Using Social Media for Risk Assessment during a Public Health Crisis. *Proceedings of the 2017 Acm Sigchi Conference on Human Factors in Computing Systems (Chi'17)*, 4520-4533.
- Gürcanlı, G., Baradan, S., & Uzun, M. (2015). Risk perception of construction equipment operators on construction sites of Turkey. *International Journal of Industrial Ergonomics*, 46, 59-68.
- Gürcanlı, G. E., & Müngen, U. (2009). An occupational safety risk analysis method at construction sites using fuzzy sets. *International Journal of Industrial Ergonomics*, 39(2), 371-387.
- Hallowell, M. (2010). Safety risk perception in construction companies in the Pacific Northwest of the USA. *Construction management and economics*, 28(4), 403-413.
- Haluik, A. (2016). Risk perception and decision making in hazard analysis: improving safety for the next generation of electrical workers. 2016 IEEE IAS Electrical Safety Workshop (ESW),
- Hansen, J., Holm, L., Frewer, L., Robinson, P., & Sandøe, P. (2003). Beyond the knowledge deficit: recent research into lay and expert attitudes to food risks. *Appetite*, 41(2), 111-121.
- Harbeck, E. L., & Glendon, A. I. (2018). Driver prototypes and behavioral willingness: Young driver risk perception and reported engagement in risky driving. *Journal of Safety Research*, 66, 195-204.
- Häring, M., Bee, N., & André, E. (2011). Creation and evaluation of emotion expression with body movement, sound and eye color for humanoid robots. 2011 RO-MAN,

- Harris, C. R., & Jenkins, M. (2006). Gender differences in risk assessment: why do women take fewer risks than men? *Judgment and Decision making*, 1(1), 48-63.
- Hartmann, K., Siegert, I., Philippou-Hübner, D., & Wendemuth, A. (2013). Emotion detection in HCI: from speech features to emotion space. *IFAC Proceedings Volumes*, 46(15), 288-295.
- Hasanzadeh, S., & De La Garza, J. M. (2020). Productivity-safety model: Debunking the myth of the productivity-safety divide through a mixed-reality residential roofing task. *Journal of Construction Engineering and Management*, 146(11), 04020124.
- Hasanzadeh, S., Esmaeili, B., & Dodd, M. D. (2018). Examining the Relationship between Construction Workers' Visual Attention and Situation Awareness under Fall and Tripping Hazard Conditions: Using Mobile Eye Tracking. *Journal of Construction Engineering and Management*, 144(7).
- Hayano, J., & Yuda, E. (2019). Pitfalls of assessment of autonomic function by heart rate variability. *Journal of Physiological Anthropology*, 38.
- He, X., Stapel, J., Wang, M., & Happee, R. (2022). Modelling perceived risk and trust in driving automation reacting to merging and braking vehicles. *Transportation research part F: traffic psychology and behaviour*, 86, 178-195.
- Heck, R. H., Thomas, S. L., & Tabata, L. N. (2013). *Multilevel and longitudinal modeling with IBM SPSS*. Routledge.
- Hegde, J., & Rokseth, B. (2020). Applications of machine learning methods for engineering risk assessment - A review. *Safety Science*, 122.
- Herrero-Fernandez, D., Macia-Guerrero, P., Silvano-Chaparro, L., Merino, L., & Jenchura, E. C. (2016). Risky behavior in young adult pedestrians: Personality

determinants, correlates with risk perception, and gender differences. *Transportation Research Part F-Traffic Psychology and Behaviour*, 36, 14-24.

Herrero-Fernández, D., Parada-Fernández, P., Oliva-Macías, M., & Jorge, R. (2020). The influence of emotional state on risk perception in pedestrians: A psychophysiological approach. *Safety Science*, 130, 104857.

Hill, A., Horswill, M. S., Whiting, J., & Watson, M. O. (2019). Computer-based hazard perception test scores are associated with the frequency of heavy braking in everyday driving. *Accident Analysis & Prevention*, 122, 207-214.

Hill, E. M., Ross, L. T., & Low, B. S. (1997). The role of future unpredictability in human risk-taking. *Human nature*, 8, 287-325.

Ho, M. C., Shaw, D., Lin, S., & Chiu, Y. C. (2008). How do disaster characteristics influence risk perception? *Risk Analysis: An International Journal*, 28(3), 635-643.

Hollnagel, E. (1998). *Cognitive reliability and error analysis method (CREAM)*. Elsevier.

Hollnagel, E. (2018). *Safety-I and safety-II: the past and future of safety management*. CRC press.

Holzman, J. B., & Bridgett, D. J. (2017). Heart rate variability indices as bio-markers of top-down self-regulatory mechanisms: A meta-analytic review. *Neuroscience & biobehavioral reviews*, 74, 233-255.

Hon, C.-Y., Randhawa, J., Lun, N., Fairclough, C., & Rothman, L. (2023). Comparison of management and workers' perception, attitudes and beliefs toward health and safety in the Ontario manufacturing sector. *Journal of Safety Research*, 84, 364-370.

Horn, M., Fovet, T., Vaiva, G., Thomas, P., Amad, A., & d'Hondt, F. (2020). Emotional response in depersonalization: A systematic review of electrodermal activity

studies. *Journal of affective disorders*, 276, 877-882.

Horswill, M. S., Garth, M., Hill, A., & Watson, M. O. (2017). The effect of performance feedback on drivers' hazard perception ability and self-ratings. *Accident Analysis & Prevention*, 101, 135-142.

Hoshi, R. A., Pastre, C. M., Vanderlei, L. C. M., & Godoy, M. F. (2013). Poincaré plot indexes of heart rate variability: relationships with other nonlinear variables. *Autonomic Neuroscience*, 177(2), 271-274.

HSC. (1993). Advisory committee on the safety of nuclear installations. In H. F. S. Group (Ed.), *Organising for Safety: Third Report of the ACSNI* (pp. 5). HMSO.

Huikuri, H. V., Mäkikallio, T. H., & Perkiömäki, J. (2003). Measurement of heart rate variability by methods based on nonlinear dynamics. *Journal of electrocardiology*, 36, 95-99.

Hutchins, G. (2018). *International Organization for Standardization, ISO 31000: 2018 enterprise risk management*. Greg Hutchins.

Ij, H. (2018). Statistics versus machine learning. *Nat Methods*, 15(4), 233.

Inouye, J. (2014). Risk perception: Theories, strategies, and next steps. *Itasca, IL: Campbell Institute National Safety Council*.

Iqbal, M. A., Ping, Q., Abid, M., Kazmi, S. M. M., & Rizwan, M. (2016). Assessing risk perceptions and attitude among cotton farmers: A case of Punjab province, Pakistan. *International Journal of Disaster Risk Reduction*, 16, 68-74.

Iqbal, M. U., Srinivasan, B., & Srinivasan, R. (2020). Dynamic assessment of control room operator's cognitive workload using Electroencephalography (EEG). *Computers & Chemical Engineering*, 141.

- Irwin, A., Tone, I. R., Sobocinska, P., Liggins, J., & Johansson, S. (2023). Thinking five or six actions ahead: Investigating the non-technical skills used within UK forestry chainsaw operations. *Safety Science*, 163.
- Janelle, C. M., Singer, R. N., & Williams, A. M. (1992). External distraction and attentional narrowing: Visual search evidence. *Journal of Sport and Exercise Psychology*, 21(1), 70-91.
- Janoff-Bulman, R. (2001). Dual-process theories in social psychology. *Contemporary Psychology-Apa Review of Books*, 46(1), 86-88.
- JASP. (2023). *JASP In* (Version 0.17.1) [free].
- Jebelli, H., Khalili, M. M., Hwang, S., & Lee, S. (2018). A supervised learning-based construction workers' stress recognition using a wearable electroencephalography (EEG) device. Construction Research Congress 2018,
- Jeelani, I., Han, K., & Albert, A. (2018). Automating and scaling personalized safety training using eye-tracking data. *Automation in Construction*, 93, 63-77.
- Jennings, M. (2020). The oil and gas industry, the competence assessment of Offshore Installation Managers (OIMs) and Control Room Operators (CROs) in emergency response, and the lack of effective assessment of underpinning technical knowledge and understanding. *Journal of Loss Prevention in the Process Industries*, 65.
- Jeon, J., & Cai, H. (2021). Classification of construction hazard-related perceptions using: Wearable electroencephalogram and virtual reality. *Automation in Construction*, 132, 103975.
- Jeon, M., & Zhang, W. (2013). Sadder but wiser? Effects of negative emotions on risk

perception, driving performance, and perceived workload. Proceedings of the Human Factors and Ergonomics Society Annual Meeting,

Ji, M., You, X. Q., Lan, J. J., & Yang, S. Y. (2011). The impact of risk tolerance, risk perception and hazardous attitude on safety operation among airline pilots in China. *Safety Science*, 49(10), 1412-1420.

Jiang, Z., Fang, D., & Zhang, M. (2015). Understanding the causation of construction workers' unsafe behaviors based on system dynamics modeling. *Journal of Management in Engineering*, 31(6), 04014099-04014099.

Jing, L., Shan, W., & Zhang, Y. (2023). Risk preference, risk perception as predictors of risky driving behaviors: the moderating effects of gender, age, and driving experience. *Journal of Transportation Safety & Security*, 15(5), 467-492.

Johnson, B. B. (1993). Advancing understanding of knowledge's role in lay risk perception. *Risk*, 4, 189.

Jorna, P. G. A. M. (1992). Spectral-Analysis of Heart-Rate and Psychological State - a Review of Its Validity as a Workload Index. *Biological Psychology*, 34(2-3), 237-257.

Kanjo, E., Younis, E. M., & Sherkat, N. (2018). Towards unravelling the relationship between on-body, environmental and emotion data using sensor information fusion approach. *Information Fusion*, 40, 18-31.

Karmakar, C. K., Gubbi, J., Khandoker, A. H., & Palaniswami, M. (2010). Analyzing temporal variability of standard descriptors of Poincaré plots. *Journal of electrocardiology*, 43(6), 719-724.

Kazemi, R., & Lee, S. C. (2023). Human Factors/Ergonomics (HFE) evaluation in the virtual reality environment: A systematic review. *International Journal of Human-Computer Interaction*, 1-17.

- Kemp, A. H., Koenig, J., & Thayer, J. F. (2017). From psychological moments to mortality: A multidisciplinary synthesis on heart rate variability spanning the continuum of time. *Neuroscience and Biobehavioral Reviews*, *83*, 547-567.
- Kemp, A. H., & Quintana, D. S. (2013). The relationship between mental and physical health: insights from the study of heart rate variability. *International journal of psychophysiology*, *89*(3), 288-296.
- Kikuta, R., Carruth, D., Ball, J., Burch, R., & Kageyama, I. (2023). Risk assessment and observation of driver with pedestrian using instantaneous heart rate and HRV. *Human Factors in Transportation*, *95*(95).
- Kim, K. H., Bang, S. W., & Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical & Biological Engineering & Computing*, *42*(3), 419-427.
- Kirwan, B. (1992). Human error identification in human reliability assessment. Part 1: Overview of approaches. *Applied ergonomics*, *23*(5), 299-318.
- Kirwan, B. (1994). *A guide to practical human reliability assessment*. CRC press.
- Kirwan, B., & James, N. (1989). *Development of a Human Reliability Assessment System for the Management of Human Error in Complex Systems*. *Reliability'89, Brighton, June 14-16*.
- Kleiger, R. E., Stein, P. K., & Bigger, J. T. (2005). Heart rate variability: Measurement and clinical utility. *Annals of Noninvasive Electrocardiology*, *10*(1), 88-101.
- Klemm, C., Hartmann, T., & Das, E. (2019). Fear-mongering or fact-driven? Illuminating the interplay of objective risk and emotion-evoking form in the response to epidemic news. *Health communication*, *34*(1), 74-83.

- Knuth, D., Kehl, D., Hulse, L., & Schmidt, S. (2014). Risk perception, experience, and objective risk: A cross - national study with European emergency survivors. *Risk Analysis*, 34(7), 1286-1298.
- Kodappully, M., Srinivasan, B., & Srinivasan, R. (2016). Towards predicting human error: Eye gaze analysis for identification of cognitive steps performed by control room operators. *Journal of Loss Prevention in the Process Industries*, 42, 35-46.
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, 84(3), 394-421.
- Krige, J., & Pestre, D. (1997). The Management of Society by Numbers. In *Science in the Twentieth Century* (1 ed., pp. 978). Routledge.
- Kumtepe, O., Akar, G. B., & Yuncu, E. (2016). Driver aggressiveness detection via multisensory data fusion. *EURASIP Journal on Image and Video Processing*, 2016, 1-16.
- La Fata, C. M., Adelfio, L., Micale, R., & La Scalia, G. (2023). Human error contribution to accidents in the manufacturing sector: A structured approach to evaluate the interdependence among performance shaping factors. *Safety Science*, 161.
- La Fata, C. M., Giallanza, A., Micale, R., & La Scalia, G. (2021). Ranking of occupational health and safety risks by a multi-criteria perspective: Inclusion of human factors and application of VIKOR. *Safety Science*, 138.
- Larkin, K. T. (2006). Psychophysiological assessment. In I. M. Hersen (Ed.), *Clinician's handbook of adult behavioral assessment* (pp. 165–186). MA: Elsevier Academic Press.
- Lawless, W. (2022). Risk determination versus risk perception: A new model of reality

for human-machine autonomy. *Informatics*,

Lee, B. G., Choi, B., Jebelli, H., & Lee, S. (2021). Assessment of construction workers' perceived risk using physiological data from wearable sensors: A machine learning approach. *Journal of Building Engineering*, 42.

Lee, G., Choi, B., Ahn, C. R., & Lee, S. (2020). Wearable Biosensor and Hotspot Analysis-Based Framework to Detect Stress Hotspots for Advancing Elderly's Mobility. *Journal of Management in Engineering*, 36(3).

Lee, G., Choi, B., Jebelli, H., & Lee, S. (2021). Assessment of construction workers' perceived risk using physiological data from wearable sensors: A machine learning approach. *Journal of Building Engineering*, 42, 102824.

Lerner, J. S., & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition & emotion*, 14(4), 473-493.

Lethaus, F., Baumann, M. R. K., Koster, F., & Lemmer, K. (2013). A comparison of selected simple supervised learning algorithms to predict driver intent based on gaze data. *Neurocomputing*, 121, 108-130.

Li, F., Chen, C. H., Lee, C. H., & Feng, S. S. (2022). Artificial intelligence-enabled non-intrusive vigilance assessment approach to reducing traffic controller's human errors. *Knowledge-Based Systems*, 239.

Li, J., Li, H., Umer, W., Wang, H. W., Xing, X. J., Zhao, S. K., & Hou, J. (2020). Identification and classification of construction equipment operators' mental fatigue using wearable eye-tracking technology. *Automation in Construction*, 109.

Li, S., Jiang, Y., Sun, C., Guo, K., & Wang, X. (2022). An Investigation on the Influence of Operation Experience on Virtual Hazard Perception Using Wearable Eye Tracking Technology. *Sensors*, 22(14), 5115.

- Li, W. (2014). Risk Assessment of Power Systems: Models, Methods, and Applications, 2nd Edition. *Risk Assessment of Power Systems: Models, Methods, and Applications, 2nd Edition*, 1-529.
- Liang, B., & Lin, Y. (2018). Using physiological and behavioral measurements in a picture-based road hazard perception experiment to classify risky and safe drivers. *Transportation research part F: traffic psychology and behaviour*, 58, 93-105.
- Liang, Y. L., Reyes, M. L., & Lee, J. D. (2007). Real-time detection of driver cognitive distraction using support vector machines. *Ieee Transactions on Intelligent Transportation Systems*, 8(2), 340-350.
- Liu, H., Li, J., Li, H., Li, H., Mao, P., & Yuan, J. (2021). Risk Perception and Coping Behavior of Construction Workers on Occupational Health Risks—A Case Study of Nanjing, China. *International Journal of Environmental Research and Public Health*, 18(13), 7040.
- Liu, J.-T., & Hsieh, C.-R. (1995). Risk perception and smoking behavior: Empirical evidence from Taiwan. *Journal of Risk and Uncertainty*, 11, 139-157.
- Liu, R. L., Cheng, W. M., Yu, Y. B., & Xu, Q. F. (2018). Human factors analysis of major coal mine accidents in China based on the HFACS-CM model and AHP method. *International Journal of Industrial Ergonomics*, 68, 270-279.
- Liu, Y., Ye, G., Xiang, Q., Yang, J., Goh, Y. M., & Gan, L. (2023). Antecedents of construction workers' safety cognition: A systematic review. *Safety Science*, 157, 105923.
- Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, 127(2), 267-286.
- Longin, E., Rautureau, G., Perez-Diaz, F., Jouvent, R., & Dubal, S. (2013). Impact of

fearful expression on danger processing: The influence of the level of trait anxiety. *Personality and Individual Differences*, 54(5), 652-657.

Lopez Vazquez, E. (2001). Risk perception interactions in stress and coping facing extreme risks. *Environmental Management and Health*, 12(2), 122-133.

Lu, J. Y., Xie, X. F., & Zhang, R. G. (2013). Focusing on appraisals: How and why anger and fear influence driving risk perception. *Journal of Safety Research*, 45, 65-73.

Lutnyk, L., Rudi, D., Schinazi, V. R., Kiefer, P., & Raubal, M. (2023). The effect of flight phase on electrodermal activity and gaze behavior: A simulator study. *Applied ergonomics*, 109, 103989.

Ma, Q., Fu, H., Xu, T., Pei, G., Chen, X., Hu, Y., & Zhu, C. (2014). The neural process of perception and evaluation for environmental hazards: evidence from event-related potentials. *Neuroreport*, 25(8), 607-611.

Ma, Y., Chowdhury, M., Sadek, A., & Jaihani, M. (2009). Real-time highway traffic condition assessment framework using vehicle–infrastructure integration (VII) with artificial intelligence (AI). *Ieee Transactions on Intelligent Transportation Systems*, 10(4), 615-627.

Ma, Y., & Ghasemzadeh, H. (2019). LabelForest: Non-parametric semi-supervised learning for activity recognition. Proceedings of the AAAI Conference on Artificial Intelligence,

MacDonald, G. (2006). Risk perception and construction safety. Proceedings of the Institution of Civil Engineers-Civil Engineering,

Magnon, V., Vallet, G. T., Benson, A., Mermillod, M., Chausse, P., Lacroix, A., Bouillon-Minois, J.-B., & Dutheil, F. (2022). Does heart rate variability predict better executive functioning? A systematic review and meta-analysis. *Cortex*.

- Makivic, B., & Bauer, P. (2017). Heart rate variability analysis in sport. *Sports Medicine*, 6, 326-331.
- Malik, M. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use: Task force of the European Society of Cardiology and the North American Society for Pacing and Electrophysiology. *Annals of Noninvasive Electrocardiology*, 1(2), 354–381.
- Man, S. S., Ng, J. Y. K., & Chan, A. H. S. (2020). A review of the risk perception of construction workers in construction safety. *Human Systems Engineering and Design II: Proceedings of the 2nd International Conference on Human Systems Engineering and Design (IHSED2019): Future Trends and Applications*, September 16-18, 2019, Universität der Bundeswehr München, Munich, Germany.
- Manar, S., Nzioka, F., & Fragouli, E. (2019). Safety & environmental risk management: borrowing from the past to enhance knowledge for the future. *Risk and Financial Management*, 1(1), 64-75.
- Manuele, F. A. (2021). Risk Assessments: Their Significance. In B. K. L. Georgi Popov, Bruce D. Hollcroft (Ed.), *Risk assessment: a practical guide to assessing operational risks* (2 ed., pp. 3).
- Marhavidas, P.-K., Koulouriotis, D., & Gemeni, V. (2011). Risk analysis and assessment methodologies in the work sites: On a review, classification and comparative study of the scientific literature of the period 2000–2009. *Journal of Loss Prevention in the Process Industries*, 24(5), 477-523.
- Marquart, G., Cabrall, C., & de Winter, J. (2015). Review of eye-related measures of drivers' mental workload. *Procedia Manufacturing*, 3, 2854-2861.
- Marris, C., Langford, I. H., & O'Riordan, T. (1998). A quantitative test of the cultural theory of risk perceptions: Comparison with the psychometric paradigm. *Risk Analysis*, 18(5), 635-647.

- Marris, C., Langford, I. H., & Riordan, T. (1996). Integrating sociological and psychological approaches to public perceptions of environmental risks: Detailed results from a questionnaire survey.
- Marshall, T. M. (2020). Risk perception and safety culture: Tools for improving the implementation of disaster risk reduction strategies. *International Journal of Disaster Risk Reduction*, 47, 101557.
- Martinez-Marquez, D., Pingali, S., Panuwatwanich, K., Stewart, R. A., & Mohamed, S. (2021). Application of eye tracking technology in aviation, maritime, and construction industries: a systematic review. *Sensors*, 21(13), 4289.
- Matos, R. (2010). Designing eye tracking experiments to measure human behavior. Merriënboer, JJG van, and Sweller, J.(2005). *Cognitive load theory and complex learning: Recent developments and future directions*. *Educational Psychology Review*, 17.
- Mbaye, S., & Kouabenan, D. R. (2013). Effects of the feeling of invulnerability and the feeling of control on motivation to participate in experience-based analysis, by type of risk. *Accident Analysis & Prevention*, 51, 310-317.
- McCourt, W. (1999). Paradigms and their development: The psychometric paradigm of personnel selection as a case study of paradigm diversity and consensus. *Organization Studies*, 20(6), 1011-1033.
- Mearns, K., & Flin, R. (1995). Risk perception and attitudes to safety by personnel in the offshore oil and gas industry: a review. *Journal of Loss Prevention in the Process Industries*, 8(5), 299-305.
- Measures, S. f. P. R. A. H. C. o. E., Boucsein, W., Fowles, D. C., Grimnes, S., Ben - Shakhar, G., Roth, W. T., Dawson, M. E., & Fillion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, 49(8), 1017-1034.

- Megias, A., Candido, A., Maldonado, A., & Catena, A. (2018). Neural correlates of risk perception as a function of risk level: An approach to the study of risk through a daily life task. *Neuropsychologia*, *119*, 464-473.
- Mesken, J., Hagenzieker, M. P., Rothengatter, T., & de Waard, D. (2007). Frequency, determinants, and consequences of different drivers' emotions: An on-the-road study using self-reports, (observed) behaviour, and physiology. *Transportation Research Part F-Traffic Psychology and Behaviour*, *10*(6), 458-475.
- Michalsen, A. (2003). Risk assessment and perception. *Injury control and safety promotion*, *10*(4), 201-204.
- Minhad, K. N., Ali, S. H. M., & Reaz, M. B. I. (2017). Happy-anger emotions classifications from electrocardiogram signal for automobile driving safety and awareness. *Journal of Transport & Health*, *7*, 75-89.
- Minkley, N., Xu, K. M., & Krell, M. (2021). Analyzing relationships between causal and assessment factors of cognitive load: associations between objective and subjective measures of cognitive load, stress, interest, and self-concept. *Frontiers in Education*,
- Mohamed, S. (2002). Safety climate in construction site environments. *Journal of Construction Engineering and Management*, *128*(5), 375-384.
- Moharreri, S., Dabanloo, N. J., & Maghooli, K. (2018). Modeling the 2D space of emotions based on the poincare plot of heart rate variability signal. *Biocybernetics and Biomedical Engineering*, *38*(4), 794-809.
- Momeni, N., Dell'Agnola, F., Arza, A., & Atienza, D. (2019). Real-time cognitive workload monitoring based on machine learning using physiological signals in rescue missions. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),

- Morioka, R. (2014). Gender difference in the health risk perception of radiation from Fukushima in Japan: the role of hegemonic masculinity. *Social Science & Medicine*, *107*, 105-112.
- Morris, J. D., & McMullen, J. S. (1994). Measuring multiple emotional responses to a single television commercial. *ACR North American Advances*.
- Morton, J., Zheleva, A., Van Acker, B. B., Durnez, W., Vanneste, P., Larmuseau, C., De Bruyne, J., Raes, A., Cornillie, F., & Saldien, J. (2022). Danger, high voltage! Using EEG and EOG measurements for cognitive overload detection in a simulated industrial context. *Applied ergonomics*, *102*, 103763.
- Moura, R., Beer, M., Patelli, E., Lewis, J., & Knoll, F. (2017). Learning from accidents: Interactions between human factors, technology and organisations as a central element to validate risk studies. *Safety Science*, *99*, 196-214.
- Mourot, L., Bouhaddi, M., Perrey, S., Cappelle, S., Henriët, M. T., Wolf, J. P., Rouillon, J. D., & Regnard, J. (2004). Decrease in heart rate variability with overtraining: assessment by the Poincaré plot analysis. *Clinical physiology and functional imaging*, *24*(1), 10-18.
- Mourot, L., Bouhaddi, M., Perrey, S., Rouillon, J.-D., & Regnard, J. (2004). Quantitative Poincaré plot analysis of heart rate variability: effect of endurance training. *European journal of applied physiology*, *91*(1), 79-87.
- Munger, S. J., Smith, R. W., & Payne, D. (1962). *An index of electronic equipment operability: data store*.
- N. Pidgeon, e. a. (1992). Risk: Analysis, Perception and Management. In T. R. Society (Ed.), *Risk perception* (pp. 89–134). The Royal Society.
- Nada, T., Nomura, M., Iga, A., Kawaguchi, R., Ochi, Y., Saito, K., Nakaya, Y., & Ito, S. (2001). Autonomic nervous function in patients with peptic ulcer studied by spectral analysis of heart rate variability. *Journal of medicine*, *32*(5-6), 333-347.

- Namian, M., Albert, A., & Feng, J. (2018). Effect of distraction on hazard recognition and safety risk perception. *Journal of Construction Engineering and Management*, 144(4), 04018008.
- Namian, M., Albert, A., Zuluaga, C. M., & Behm, M. (2016). Role of safety training: Impact on hazard recognition and safety risk perception. *Journal of Construction Engineering and Management*, 142(12), 04016073-04016071-04016073-04016010.
- Nasirzadeh, F., Mir, M., Hussain, S., Tayarani Darbandy, M., Khosravi, A., Nahavandi, S., & Aisbett, B. (2020). Physical fatigue detection using entropy analysis of heart rate signals. *Sustainability*, 12(7), 2714.
- Network, C. I. (2023). *ErgoLAB: HRV Analysis Module*.
- Ng, S. L. (2023). The role of risk perception, prior experience, and sociodemographics in disaster preparedness and emergency response toward typhoons in Hong Kong. *Natural Hazards*, 116(1), 905-936.
- Nicholas, B. (2006). Risk perception and safety management systems in the global maritime industry. *Policy and Practice in Health and Safety*, 4(2), 59-75.
- Niezgoda, M., Tarnowski, A., Kruszewski, M., & Kamiński, T. (2015). Towards testing auditory–vocal interfaces and detecting distraction while driving: A comparison of eye-movement measures in the assessment of cognitive workload. *Transportation research part F: traffic psychology and behaviour*, 32, 23-34.
- Nishiyama, T., Sugeno, J., Matsumoto, T., Iwase, S., & Mano, T. (2001). Irregular activation of individual sweat glands in human sole observed by a videomicroscopy. *Autonomic Neuroscience*, 88(1-2), 117-126.

- Nitzan, M., Babchenko, A., Khanokh, B., & Landau, D. (1998). The variability of the photoplethysmographic signal - a potential method for the evaluation of the autonomic nervous system. *Physiological Measurement*, 19(1), 93-102.
- Nunan, D., Sandercock, G. R., & Brodie, D. A. (2010). A quantitative systematic review of normal values for short - term heart rate variability in healthy adults. *Pacing and clinical electrophysiology*, 33(11), 1407-1417.
- Okrent, D. (1998). Risk perception and risk management: on knowledge, resource allocation and equity. *Reliability Engineering & System Safety*, 59(1), 17-25.
- Orlandi, L., & Brooks, B. (2018). Measuring mental workload and physiological reactions in marine pilots: Building bridges towards redlines of performance. *Applied ergonomics*, 69, 74-92.
- Ozel, P., Akan, A., & Yilmaz, B. (2019). Synchrosqueezing transform based feature extraction from EEG signals for emotional state prediction. *Biomedical signal processing and control*, 52, 152-161.
- P.K. Marhavilas, P. K. M., V. Gemeni. (2011). Risk analysis and assessment methodologies in the work sities: On a review, classification and comparative study of the scientific literature of the period 2000e2009 *Journal of Loss Prevention in the Process Industries*, 24, 477-523.
- Pakdamanian, E., Feng, L., & Kim, I. (2018). The effect of whole-body haptic feedback on driver's perception in negotiating a curve. Proceedings of the Human Factors and Ergonomics Society Annual Meeting,
- Pandit, B., Albert, A., Patil, Y., & Al-Bayati, A. J. (2019). Impact of safety climate on hazard recognition and safety risk perception. *Safety Science*, 113, 44-53.
- Panicker, S. S., & Gayathri, P. (2019). A survey of machine learning techniques in physiology based mental stress detection systems. *Biocybernetics and Biomedical Engineering*, 39(2), 444-469.

- Park, B. (2009). Psychophysiology as a Tool for HCI Research: Promises and Pitfalls. *Human-Computer Interaction, Pt I, 5610*, 141-148.
- Park, S., Park, C. Y., Lee, C., Han, S. H., Yun, S., & Lee, D.-E. (2022). Exploring inattentive blindness in failure of safety risk perception: Focusing on safety knowledge in construction industry. *Safety Science, 145*, 105518.
- Parrott, W. G. (2017). Role of emotions in risk perception. *Consumer perception of product risks and benefits*, 221-232.
- Pasman, H. J., Knegtering, B., & Rogers, W. (2013). A holistic approach to control process safety risks: Possible ways forward. *Reliability Engineering & System Safety, 117*, 21-29.
- Pate-Cornell, M. E., & Murphy, D. M. (1996). Human and management factors in probabilistic risk analysis: the SAM approach and observations from recent applications. *Reliability Engineering & System Safety, 53(2)*, 115-126.
- Patel, D., & Jha, K. (2015). Neural network model for the prediction of safe work behavior in construction projects. *Journal of Construction Engineering and Management, 141(1)*, 04014066.
- Paton, D., Johnston, D., Bebbington, M. S., Lai, C.-D., & Houghton, B. F. (2000). Direct and vicarious experience of volcanic hazards: implications for risk perception and adjustment adoption. *Australian Journal of Emergency Management, The, 15(4)*, 58-63.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research, 12*, 2825-2830.

- Penttilä, J., Helminen, A., Jartti, T., Kuusela, T., Huikuri, H. V., Tulppo, M. P., Coffeng, R., & Scheinin, H. (2001). Time domain, geometrical and frequency domain analysis of cardiac vagal outflow: effects of various respiratory patterns. *Clinical physiology*, 21(3), 365-376.
- Perello-March, J. R., Burns, C. G., Birrell, S. A., Woodman, R., & Elliott, M. T. (2022). Physiological Measures of Risk Perception in Highly Automated Driving. *Ieee Transactions on Intelligent Transportation Systems*, 23(5), 4811-4822.
- Perlman, A., Sacks, R., & Barak, R. (2014). Hazard recognition and risk perception in construction. *Safety Science*, 64, 22-31.
- Picard, R. W., Fedor, S., & Ayzenberg, Y. (2016). Multiple arousal theory and daily-life electrodermal activity asymmetry. *Emotion review*, 8(1), 62-75.
- Pidgeon, N. (1998). Risk assessment, risk values and the social science programme: why we do need risk perception research. *Reliability Engineering & System Safety*, 59(1), 5-15.
- Pidgeon, N., Hood, C., Jones, D., Turner, B., & Gibson, R. (1992). Risk perception. In *Risk: Analysis, perception and management* (pp. 89-134). The Royal Society.
- Ping, P., Sheng, Y., Qin, W., Miyajima, C., & Takeda, K. (2018). Modeling driver risk perception on city roads using deep learning. *IEEE Access*, 6, 68850-68866.
- Platania, M., Platania, S., & Santisi, G. (2016). Entertainment marketing, experiential consumption and consumer behavior: the determinant of choice of wine in the store. *Wine Economics and Policy*, 5(2), 87-95.
- Poh, M.-Z., Swenson, N. C., & Picard, R. W. (2010). A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *Ieee Transactions on Biomedical Engineering*, 57(5), 1243-1252.

- Porges, S. W. (1995). Orienting in a defensive world: Mammalian modifications of our evolutionary heritage. A polyvagal theory. *Psychophysiology*, 32(4), 301-318.
- Porges, S. W. (1997). Emotion: An evolutionary by-product of the neural regulation of the autonomic nervous system. *Annals of the New York Academy of Sciences-Paper Edition*, 807, 62-77.
- Porges, S. W. (2001). The polyvagal theory: phylogenetic substrates of a social nervous system. *International journal of psychophysiology*, 42(2), 123-146.
- Porges, S. W. (2003a). The polyvagal theory: Phylogenetic contributions to social behavior. *Physiology & behavior*, 79(3), 503-513.
- Porges, S. W. (2003b). Social engagement and attachment: A phylogenetic perspective. *Annals of the New York Academy of Sciences*, 1008(1), 31-47.
- Porges, S. W. (2007). The polyvagal perspective. *Biological Psychology*, 74(2), 116-143.
- Porges, S. W. (2009). The polyvagal theory: new insights into adaptive reactions of the autonomic nervous system. *Cleveland Clinic journal of medicine*, 76(Suppl 2), S86.
- Powers, B. J., Oddone, E. Z., Grubber, J. M., Olsen, M. K., & Bosworth, H. B. (2008). Perceived and actual stroke risk among men with hypertension. *The Journal of Clinical Hypertension*, 10(4), 287-294.
- Pradhan, A. K., Fisher, D. L., & Pollatsek, A. (2006). Risk perception training for novice drivers: evaluating duration of effects of training on a driving simulator. *Transportation Research Record*, 1969(1), 58-64.

- Prell, R., Opatz, O., Merati, G., Gesche, B., Gunga, H.-C., & Maggioni, M. A. (2020). Heart rate variability, risk-taking behavior and resilience in firefighters during a simulated extinguish-fire task. *Frontiers in Physiology*, *11*, 482.
- Provan, D. J., Woods, D. D., Dekker, S. W., & Rae, A. J. (2020). Safety II professionals: How resilience engineering can transform safety practice. *Reliability Engineering & System Safety*, *195*, 106740.
- Rajendra Acharya, U., Paul Joseph, K., Kannathal, N., Lim, C. M., & Suri, J. S. (2006). Heart rate variability: a review. *Medical and biological engineering and computing*, *44*, 1031-1051.
- Rawal, K., & Ahmad, A. (2021). Feature selection for electrical demand forecasting and analysis of pearson coefficient. 2021 IEEE 4th International Electrical and Energy Conference (CIEEC),
- Reader, T. W., & O'Connor, P. (2014). The Deepwater Horizon explosion: non-technical skills, safety culture, and system complexity. *Journal of Risk Research*, *17*(3), 405-424.
- Rechard, R. P. (1999). Historical relationship between performance assessment for radioactive waste disposal and other types of risk assessment. *Risk Analysis*, *19*(5), 763-807.
- Rechard, R. P. (2000). Historical background on performance assessment for the Waste Isolation Pilot Plant. *Reliability Engineering & System Safety*, *69*(1-3), 5-46.
- Rendon-Velez, E., van Leeuwen, P. M., Happee, R., Horvath, I., van der Vegte, W. F., & de Winter, J. C. F. (2016). The effects of time pressure on driver performance and physiological activity: A driving simulator study. *Transportation Research Part F-Traffic Psychology and Behaviour*, *41*, 150-169.
- Renn, O. (1992). Concepts of risk: a classification. In S. G. Krimsky, D (Ed.), *Social Theories of Risk* (pp. 53-79). CT: Praeger.

- Renn, O. (1998). The role of risk perception for risk management. *Reliability Engineering & System Safety*, 59(1), 49-62.
- Riad, J. K., Norris, F. H., & Ruback, R. B. (1999). Predicting evacuation in two major disasters: Risk perception, social influence, and access to resources 1. *Journal of applied social Psychology*, 29(5), 918-934.
- Rincon, J. A., de la Prieta, F., Zanardini, D., Julian, V., & Carrascosa, C. (2017). Influencing over people with a social emotional model. *Neurocomputing*, 231, 47-54.
- Rosi, G., Vignali, G., & Bottani, E. (2018). A conceptual framework for the selection of an "Industry 4.0" application to enhance the operators' safety: the case of an aseptic bottling line. *2018 Ieee International Conference on Engineering, Technology and Innovation (Ice/Itmc)*.
- Rowe, D. W., Sibert, J., & Irwin, D. (1998). *Heart rate variability: Indicator of user state as an aid to human-computer interaction* Proceedings of the SIGCHI conference on Human factors in computing systems,
- Rubaltelli, E., Scrimin, S., Moscardino, U., Priolo, G., & Buodo, G. (2018). Media exposure to terrorism and people's risk perception: The role of environmental sensitivity and psychophysiological response to stress. *British Journal of Psychology*, 109(4), 656-673.
- Rundmo, T. (1992a). Risk perception and safety on offshore petroleum platforms—Part I: Perception of risk. *Safety Science*, 15(1), 39-52.
- Rundmo, T. (1992b). Risk perception and safety on offshore petroleum platforms—Part II: Perceived risk, job stress and accidents. *Safety Science*, 15(1), 53-68.

- Rundmo, T. (1995). Perceived risk, safety status, and job stress among injured and noninjured employees on offshore petroleum installations. *Journal of Safety Research*, 26(2), 87-97.
- Rundmo, T. (1996). Associations between risk perception and safety. *Safety Science*, 24(3), 197-209.
- Rundmo, T. (2000). Safety climate, attitudes and risk perception in Norsk Hydro. *Safety Science*, 34(1-3), 47-59.
- Saboul, D., Pialoux, V., & Hautier, C. (2013). The impact of breathing on HRV measurements: Implications for the longitudinal follow-up of athletes. *European journal of sport science*, 13(5), 534-542.
- Sandman, P. M. (1989). Hazard versus outrage in the public perception of risk. In *Effective risk communication: The role and responsibility of government and nongovernment organizations* (pp. 45-49). Springer.
- Sarker, I. H., Abushark, Y. B., Alsolami, F., & Khan, A. I. (2020). Intrudtree: a machine learning based cyber security intrusion detection model. *Symmetry*, 12(5), 754.
- Schmidt, E., Decke, R., Rasshofer, R., & Bullinger, A. C. (2017). Psychophysiological responses to short-term cooling during a simulated monotonous driving task. *Applied ergonomics*, 62, 9-18.
- Schmidt, S., & von Stülpnagel, R. (2018). *Risk perception and gaze behavior during urban cycling—a field study* Eye Tracking for Spatial Research, Proceedings of the 3rd International Workshop,
- Schnittker, R. (2012). *Electrodermal activity of novice drivers during driving simulator training—an explorative study* University of Twente].

scikit-learn. (2023). *Feature importances with a forest of trees*.

Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in public health*, 258.

Shaffer, F., McCraty, R., & Zerr, C. L. (2014). A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. *Frontiers in psychology*, 5, 1040.

Shayesteh, S., Ojha, A., Liu, Y. Z., & Jebelli, H. (2023). Human-robot teaming in construction: Evaluative safety training through the integration of immersive technologies and wearable physiological sensing. *Safety Science*, 159.

Shin, M., Lee, H.-S., Park, M., Moon, M., & Han, S. (2014). A system dynamics approach for modeling construction workers' safety attitudes and behaviors. *Accident Analysis & Prevention*, 68, 95-105.

Shou, Y., & Olney, J. (2021). Attitudes toward risk and uncertainty: The role of subjective knowledge and affect. *Journal of Behavioral Decision Making*, 34(3), 393-404.

Siegrist, M., & Árvai, J. (2020). Risk perception: Reflections on 40 years of research. *Risk Analysis*, 40(S1), 2191-2206.

Siegrist, M., Keller, C., & Kiers, H. A. (2005). A new look at the psychometric paradigm of perception of hazards. *Risk Analysis: An International Journal*, 25(1), 211-222.

Singleton, W. T., & Hovden, J. (1987). *Risk and decisions*. John Wiley & Sons Incorporated.

Sitkin, S. B., & Pablo, A. L. (1992). Reconceptualizing the determinants of risk behavior. *Academy of management review*, 17(1), 9-38.

- Sitkin, S. B., & Weingart, L. R. (1995). Determinants of risky decision-making behavior: A test of the mediating role of risk perceptions and propensity. *Academy of management Journal*, 38(6), 1573-1592.
- Sjöberg, L. (2006). Myths of the psychometric paradigm and how they can misinform risk communication. Risk Perception and Communication Consultation Technical Meeting, organized by World Health Organization, Regional Office for Europe, Venice, Italy,
- Sjöberg, L. (2007). Emotions and risk perception. *Risk Management*, 9(4), 223-237.
- Sjöberg, L., Moen, B.-E., & Rundmo, T. (2004). *Explaining risk perception. An evaluation of the psychometric paradigm in risk perception research* (Vol. 84). Norwegian University of Science and Technology, Department of Psychology.
- Skjong, R., & Wentworth, B. H. (2001). Expert judgment and risk perception. ISOPE International Ocean and Polar Engineering Conference,
- Slovic, P. (1987). Perception of risk. *Science*, 236, 208-285.
- Slovic, P. (1988). Risk perception. *Carcinogen risk assessment*, 171-181.
- Slovic, P. (1990). *Perception of risk: Reflections on the psychometric paradigm* (D. Golding & S. Krimsky, Eds.). Praeger.
- Slovic, P. (1992). Perception of risk: Reflections on the psychometric paradigm. In.
- Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2004). Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk*

- Slovic, P., Fischhoff, B., & Lichtenstein, S. (1982). Why study risk perception? *Risk Analysis*, 2(2), 83-93.
- Slovic, P., Fischhoff, B., & Lichtenstein, S. (1985). Characterizing perceived risk. *Perilous progress: Managing the hazards of technology*, 91-125.
- Slovic, P., & Peters, E. (2006). Risk perception and affect. *Current directions in psychological science*, 15(6), 322-325.
- Slovic, P., & Weber, E. U. (2013). Perception of risk posed by extreme events. *Regulation of Toxic Substances and Hazardous Waste (2nd edition)*(Applegate, Gabba, Laitos, and Sachs, Editors), Foundation Press, Forthcoming.
- Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, 4(2), 108-131.
- Sohail, A., Cheema, M. A., Ali, M. E., Toosi, A. N., & Rakha, H. A. (2023). Data-driven approaches for road safety: a comprehensive systematic literature review. *Safety Science*, 158, 105949.
- SPSSPRO. (2021). *Scientific Platform Serving for Statistics Professional 2021*. In *SPSSPRO* (Version 1.0.11) [Online Application Software].
- Srinivasan, R., Srinivasan, B., Iqbal, M. U., Nemet, A., & Kravanja, Z. (2019). Recent developments towards enhancing process safety: Inherent safety and cognitive engineering. *Computers & Chemical Engineering*, 128, 364-383.
- Stamatelatos, M., Vesely, W., Dugan, J., Fragola, J., Minarick, J., & Railsback, J. (2002). *Fault tree handbook with aerospace applications*. NASA Office of Safety and

Mission Assurance.

Steg, L., & Sievers, I. (2000). Cultural theory and individual perceptions of environmental risks. *Environment and behavior*, 32(2), 250-269.

Stülpnagel, R. v., Petinaud, C., & Lißner, S. (2022). Crash risk and subjective risk perception during urban cycling: Accounting for cycling volume. *Accident Analysis & Prevention*, 164, 106470.

Suraji, A., Duff, A. R., & Peckitt, S. J. (2001). Development of causal model of construction accident causation. *Journal of Construction Engineering and Management*, 127(4), 337-344.

Swain, A. D., & Guttman, H. E. (1983). *Handbook of human-reliability analysis with emphasis on nuclear power plant applications. Final report.*

Swuste, P., van Gulijk, C., Zwaard, W., & Oostendorp, Y. (2014). Occupational safety theories, models and metaphors in the three decades since World War II, in the United States, Britain and the Netherlands: A literature review. *Safety Science*, 62, 16-27.

Sztajzel, J. (2004). Heart rate variability: a noninvasive electrocardiographic method to measure the autonomic nervous system. *Swiss medical weekly*, 134(3536), 514-522.

Tagliabue, M., & Sarlo, M. (2015). Affective components in training to ride safely using a moped simulator. *Transportation Research Part F-Traffic Psychology and Behaviour*, 35, 132-138.

Tagliabue, M., Sarlo, M., & Gianfranchi, E. (2019). How can on-road hazard perception and anticipation be improved? Evidence from the body. *Frontiers in psychology*, 10, 167.

- Tang, J., Yang, S., Liu, Y., Yao, K., & Wang, G. (2022). Typhoon Risk Perception: A Case Study of Typhoon Lekima in China. *International Journal of Disaster Risk Science*, 13(2), 261-274.
- Tanida, K., Paolini, M., Poppel, E., & Silveira, S. (2018). Safety feelings and anticipatory control: An fMRI study on safety and risk perception. *Transportation Research Part F-Traffic Psychology and Behaviour*, 57, 108-114.
- Taylor, D. (1964). DRIVERS' GALVANIC SKIN RESPONSE AND THE RISK OF ACCIDENT. *Ergonomics*, 7(4), 439-451.
- Taylor, S. E. (2006). Tend and befriend: Biobehavioral bases of affiliation under stress. *Current directions in psychological science*, 15(6), 273-277.
- Taylor, W. D., & Snyder, L. A. (2017). The influence of risk perception on safety: A laboratory study. *Safety Science*, 95, 116-124.
- Tixier, A. J.-P., Hallowell, M. R., Albert, A., van Boven, L., & Kleiner, B. M. (2014). Psychological antecedents of risk-taking behavior in construction. *Journal of Construction Engineering and Management*, 140(11), 04014052.
- Tixier, A. J. P., Hallowell, M. R., Rajagopalan, B., & Bowman, D. (2017). Construction Safety Clash Detection: Identifying Safety Incompatibilities among Fundamental Attributes using Data Mining. *Automation in Construction*, 74, 39-54.
- Tjolleng, A., Jung, K., Hong, W., Lee, W., Lee, B., You, H., Son, J., & Park, S. (2017). Classification of a Driver's cognitive workload levels using artificial neural network on ECG signals. *Applied ergonomics*, 59, 326-332.
- Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology*, 49(11), 1225-1231.

- Tulppo, M. P., Makikallio, T. H., Takala, T., Seppanen, T., & Huikuri, H. V. (1996). Quantitative beat-to-beat analysis of heart rate dynamics during exercise. *American journal of physiology-heart and circulatory physiology*, 271(1), 244-252.
- Van der Velde, F. W., Hooykaas, C., & Van der Joop, P. (1992). Risk perception and behavior: Pessimism, realism, and optimism about AIDS-related health behavior. *Psychology and Health*, 6(1-2), 23-38.
- Västfjäll, D., Peters, E., & Slovic, P. (2008). Affect, risk perception and future optimism after the tsunami disaster. *Judgment and Decision making*, 3(1), 64-72.
- Veland, H., & Aven, T. (2013). Risk communication in the light of different risk perspectives. *Reliability Engineering & System Safety*, 110, 34-40.
- Verkijika, S. F., & De Wet, L. (2019). Understanding word-of-mouth (WOM) intentions of mobile app users: The role of simplicity and emotions during the first interaction. *Telematics and Informatics*, 41, 218-228.
- Villa, V., Paltrinieri, N., Khan, F., & Cozzani, V. (2016). Towards dynamic risk analysis: A review of the risk assessment approach and its limitations in the chemical process industry. *Safety Science*, 89, 77-93.
- Wachinger, G., Renn, O., Begg, C., & Kuhlicke, C. (2013). The risk perception paradox—implications for governance and communication of natural hazards. *Risk Analysis*, 33(6), 1049-1065.
- Wade, N. J., & Tatler, B. W. (2009). Did Javal measure eye movements during reading? *Journal of Eye Movement Research*, 2(5).
- Wahl, A., Kongsvik, T., & Antonsen, S. (2020). Balancing Safety I and Safety II: Learning

to manage performance variability at sea using simulator-based training. *Reliability Engineering & System Safety*, 195, 106698.

Wang, J., Zou, P. X., & Li, P. P. (2016). Critical factors and paths influencing construction workers' safety risk tolerances. *Accident Analysis & Prevention*, 93, 267-279.

Wang, W., Li, Z., Wang, Y., & Chen, F. (2013). Indexing cognitive workload based on pupillary response under luminance and emotional changes. Proceedings of the 2013 international conference on Intelligent user interfaces,

Wang, X. H., Huang, X. F., Luo, Y., Pei, J. J., & Xu, M. (2018). Improving Workplace Hazard Identification Performance Using Data Mining. *Journal of Construction Engineering and Management*, 144(8).

Waring, A. (2015). Managerial and non-technical factors in the development of human-created disasters: A review and research agenda. *Safety Science*, 79, 254-267.

Watling, C. N., Hasan, M. M., & Larue, G. S. (2021). Sensitivity and specificity of the driver sleepiness detection methods using physiological signals: A systematic review. *Accident Analysis and Prevention*, 150.

Watts, G., & Quimby, A. (1979). *Design and validation of a driving simulator* (Report LR, Issue. T. a. R. R. Laboratory.

Weber, E. U. (2017). Understanding public risk perception and responses to changes in perceived risk. *Policy shock: Recalibrating risk and regulation after oil spills, nuclear accidents and financial crises*, 58-81.

Weber, E. U., Blais, A. R., & Betz, N. E. (2002). A domain - specific risk - attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, 15(4), 263-290.

- Wehrwein, E. A., Orer, H. S., & Barman, S. M. (2016). Overview of the anatomy, physiology, and pharmacology of the autonomic nervous system. *regulation*, 37(69), 125.
- Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). *Engineering psychology and human performance*. Psychology Press.
- Widyanti, A., Muslim, K., & Sotalaksana, I. Z. (2017). The sensitivity of Galvanic Skin Response for assessing mental workload in Indonesia. *Work*, 56(1), 111-117.
- Wilde, G. (2014). *Target Risk 3 Risk Homeostasis in Everyday Life*. Dublin.
- Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology*, 12(1), 3-18.
- Wilson, R. S., Zwickle, A., & Walpole, H. (2019). Developing a broadly applicable measure of risk perception. *Risk Analysis*, 39(4), 777-791.
- Windsor, T. D., Anstey, K. J., & Walker, J. G. (2008). Ability perceptions, perceived control, and risk avoidance among male and female older drivers. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 63(2), P75-P83.
- Wong, J. C. S., & Yang, J. Z. (2023). Risk perception of the COVID-19 vaccines: revisiting the psychometric paradigm. *Journal of Risk Research*, 26(6), 697-709.
- Woodcock, B., & Au, Z. (2013). Human factors issues in the management of emergency response at high hazard installations. *Journal of Loss Prevention in the Process Industries*, 26(3), 547-557.
- Wu, D., & Cui, Y. (2018). Disaster early warning and damage assessment analysis using

social media data and geo-location information. *Decision support systems*, 111, 48-59.

Wundt, W. (1922). Vorlesung über die Menschen-und Tierseele. *Siebente und Achte Auflage*.

Xia, N., Wang, X., Griffin, M. A., Wu, C., & Liu, B. (2017). Do we see how they perceive risk? An integrated analysis of risk perception and its effect on workplace safety behavior. *Accident Analysis & Prevention*, 106, 234-242.

Xiao, Z., Ponnambalam, L., Fu, X. J., & Zhang, W. B. (2017). Maritime Traffic Probabilistic Forecasting Based on Vessels' Waterway Patterns and Motion Behaviors. *Ieee Transactions on Intelligent Transportation Systems*, 18(11), 3122-3134.

Xie, Q., & Xue, Y. (2022). The prediction of public risk perception by internal characteristics and external environment: Machine learning on big data. *International Journal of Environmental Research and Public Health*, 19(15), 9545.

Xinran Zhang, & Yan, X. (2023). Predicting collision cases at unsignalized intersections using EEG metrics and driving simulator platform. *Accident Analysis and Prevention*, 180.

Xiong, C. a., & Guo, F. (2021). Psychological Characteristics of Elderly Drivers in Risk Scenarios Based on Heart Rate Variability. In *CICTP 2021* (pp. 985-995).

Xiong, R. X., Wang, Y. Y., Tang, P. B., Cooke, N. J., Ligda, S. V., Lieber, C. S., & Liu, Y. M. (2023). Predicting separation errors of air traffic controllers through integrated sequence analysis of multimodal behaviour indicators. *Advanced Engineering Informatics*, 55.

Xu, J., Anders, S., Pruttianan, A., France, D., Lau, N., Adams, J. A., & Weinger, M. B. (2018). Human performance measures for the evaluation of process control

human-system interfaces in high-fidelity simulations. *Applied ergonomics*, 73, 151-165.

Xu, Z., Karwowski, W., Çakıt, E., Reineman-Jones, L., Murata, A., Aljuaid, A., Sapkota, N., & Hancock, P. (2023). Nonlinear dynamics of EEG responses to unmanned vehicle visual detection with different levels of task difficulty. *Applied ergonomics*, 111, 104045.

Yan, S., Wei, Y., & Tran, C. C. (2019). Evaluation and prediction mental workload in user interface of maritime operations using eye response. *International Journal of Industrial Ergonomics*, 71, 117-127.

Yang, J. Z., & Chu, H. (2018). Who is afraid of the Ebola outbreak? The influence of discrete emotions on risk perception. *Journal of Risk Research*, 21(7), 834-853.

Yang, Z., & Liu, G. (2014). An entropy measure of emotional arousal via skin conductance response. *Journal of Fiber Bioengineering and Informatics*, 7(1), 67-80.

Yu, L., & Liu, H. (2004). Efficient feature selection via analysis of relevance and redundancy. *The Journal of Machine Learning Research*, 5, 1205-1224.

Zermane, A., Tohir, M. Z. M., Baharudin, M. R., & Yusoff, H. M. (2022). Risk assessment of fatal accidents due to work at heights activities using fault tree analysis: Case study in Malaysia. *Safety Science*, 151.

Zhai, J., & Barreto, A. (2006). *Stress Recognition Using Non-invasive Technology* FLAIRS conference,

Zhang, M., Jipp, M., & Ihme, K. (2022). The novelty appraisal of the feeling of risk in vehicles. *International Journal of Environmental Research and Public Health*, 19(21), 14259.

- Zhang, P., Lingard, H., Blismas, N., Wakefield, R., & Kleiner, B. (2015). Work-health and safety-risk perceptions of construction-industry stakeholders using photograph-based Q methodology. *Journal of Construction Engineering and Management*, *141*(5), 04014093.
- Zhang, X., & Yan, X. (2023). Predicting collision cases at unsignalized intersections using EEG metrics and driving simulator platform. *Accident Analysis & Prevention*, *180*, 106910.
- Zhou, J., Yu, K., Chen, F., Wang, Y., & Arshad, S. Z. (2018). Multimodal behavioral and physiological signals as indicators of cognitive load. In *The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition-Volume 2* (pp. 287-329).
- Zhou, Z.-H. (2012). *Ensemble methods: foundations and algorithms*. CRC press.
- Zhu, R., Wang, Z., Ma, X., & You, X. (2022). High expectancy influences the role of cognitive load in inattentive deafness during landing decision-making. *Applied ergonomics*, *99*, 103629.
- Zimmer, P., Buttler, B., Halbeisen, G., Walther, E., & Domes, G. (2019). Virtually stressed? A refined virtual reality adaptation of the Trier Social Stress Test (TSST) induces robust endocrine responses. *Psychoneuroendocrinology*, *101*, 186-192.
- Zio, E. (2018). The future of risk assessment. *Reliability Engineering & System Safety*, *177*, 176-190.
- Zsido, A. N., Csokasi, K., Vincze, O., & Coelho, C. M. (2020). The emergency reaction questionnaire - First steps towards a new method. *International Journal of Disaster Risk Reduction*, *49*.