NOVA IMS Information Management School

MGI

Master Degree Program in

Information Management

How to Detect and Mitigate Burnout in Organizations by leveraging Data Analytics

Francisco Sarreira Nunes de Oliveira

Master's Thesis

presented as partial requirement for obtaining a Master’s Degree in Information Management

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

How to Detect and Mitigate Burnout in Organizations by leveraging Data Analytics

by

Francisco Sarreira Nunes de Oliveira

Master Thesis presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Information Systems Management

Supervised by

Prof. Vítor Duarte dos Santos, PhD, NOVA Information Management School

July, 2025

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honour from the NOVA Information Management School.

Lisbon, July 2025

Francisco Sarreira Nunes de Oliveira

ACKNOWLEDGEMENTS

Writing this thesis was, at many times, a real challenge. Over the course of these two years of my master's degree, many things changed in both my personal and professional life. These changes often required resilience, patience and hope for things to finally start falling into place. Working on this topic was especially meaningful to me because, in many ways, I experienced it firsthand. I went through professional changes, major personal shifts and even during the development of this thesis, there were moments when I felt lost and unmotivated.

Still, I was never truly alone. I felt tremendous support from my supervisor, Professor Vítor Duarte dos Santos, who, even during my lowest moments, never stopped believing in me. He continued to encourage me, face challenges with me and helped me overcome them. He has my deepest and most sincere gratitude.

I also want to thank my classmates, who showed great patience, offered their help and brought laughter into some of the most difficult times. Without a doubt, they made my academic journey much more enjoyable. Thank you for that.

A very special thank you to my family and friends, who were truly a pillar of strength for me. Whether it was to laugh, cry or simply listen to my frustrations, you were always there to support me. From the bottom of my heart, thank you.

Lastly, a message to myself. It wasn't easy. Not just because of the work, but because of what you were carrying inside while doing it. There were days you wanted to give up, days you felt like the ground beneath you had disappeared and yet, you kept going.
This wasn't just a thesis. It was learning to trust yourself, to accept the chaos, to find strength where you thought there was none left. It was growing on the inside, even when everything on the outside seemed the same. You managed to get through it, with flaws, but with courage too. It was a tough experience but incredibly rewarding. I'm proud of you.

ABSTRACT

Burnout has become a critical concern for organisations with significant implications for employee well-being and organisational productivity, yet data-driven approaches to its early detection and mitigation remain underexplored. This study investigates how data analytics can be leveraged to identify early signs of stress and burnout in the workplace and inform effective interventions. The research includes a systematic literature review (PRISMA) with an empirical analysis of an open dataset of 831 healthcare workers surveyed during the COVID-19 pandemic. Key variables such as burnout levels (from the Maslach Burnout Inventory), working hours, shift type, resilience and COVID-19-related anxiety were analysed using statistical techniques and supervised machine learning models (Logistic Regression, Decision Tree and Random Forest). The results reveal that burnout was alarmingly prevalent in the sample (97.6% of respondents in moderate or high burnout). Resilience emerged as a protective factor showing a moderate negative correlation (r = -0.38) with exhaustion, while fear of COVID-19 showed a positive correlation, suggesting personal psychological resources can buffer stress, whereas pandemic-related stressors make it worse. Extended working hours (>60 h/week) and night/rotating shifts were associated with higher exhaustion. Using these insights, the study proposes a data-driven intervention framework: for instance, employees with high burnout and low resilience can be targeted with workload adjustments, resilience training or professional support. The results show how analytics may be used to identify employees who are at risk and inform customized burnout prevention plans. By integrating organisational data and employee well-being metrics, organisations can proactively address burnout, improving both employee health and workplace productivity. This research contributes to the field of information management by showing how quantitative data can drive evidence-based HR and wellness interventions and It establishes the foundation for more flexible, data-informed approaches to safeguarding employee well-being.

KEYWORDS

Data Analytics; Stress; Burnout; Employee Well-being; Data-Driven Strategies

Sustainable Development Goals (SDG):

INDEX

1 Introduction 10
1.1 Context and Problem definition 10
1.2 Objectives 11
1.3 Study Relevance and Importance 12
2 Literature Review 14
2.1 Burnout 14
2.1.1 Concepts 14
2.1.2 Measuring 14
2.2 Systematic Literature Review on Burnout & Data Science 15
2.2.1 Inclusion and Exclusion Criteria 15
2.2.2 Study Selection 16
2.2.3 Results Analysis 17
3 Methodology 30
3.1 Research Design 30
3.2 Hypotheses 30
3.3 Data Collection 31
3.4 Analytical Methods 32
3.5 Ethical Considerations 32
3.6 Limitations 33
4 Empirical Study 34
4.1 Business Understanding 34
4.2 Data Collection 34
4.3 Data understanding 35
4.4 Data Preparation 35
4.5 Data Reduction 35
4.6 Execution 36
5 Results and Discussion 37
6 Conclusions and Future Research 46
6.1 Theoretical and practical implications 46
6.2 Limitations 47
6.3 Future Research 47
Bibliographical References 49

Appendix A - Questionnaire 51
Appendix B: Coding for the figures 56
Appendix C: Generative AI Assistance 66

LIST OF FIGURES

Figure 1: Study selection flowchart (PRISMA) 17
Figure 2 Occupational Burnout Levels Distribution 37
Figure 3 Age Distribution of Healthcare Workers 38
Figure 4 Burnout Levels by Gender 39
Figure 5 Burnout vs Weekly Working Hours 40
Figure 6 Correlation Matrix of Key Variables 41
Figure 7 Comprehensive Analysis: Resilience vs Burnout 42
Figure 8 Intervention Targets Based on Burnout and Resilience Profiles 44

LIST OF TABLES

Table 1 Correlation between the measuring tools 15
Table 2 Prima table..... 18
Table 3 Connection between studies and hypothesis 31
Table 4 Clinical Action Framework..... 45

LIST OF ABBREVIATIONS AND ACRONYMS

APA American Psychological Association

CDC Centers for Disease Control and Prevention

DP Depersonalization

EE Emotional Exhaustion

FCV-19S Fear of COVID-19 Scale

ICD International Classification of Diseases

IDE Integrated development environment

MBI Maslach Burnout Inventory

MBI-ES Maslach Burnout Inventory – Educators Survey

MBI-GS Maslach Burnout Inventory – General Survey

MBI-HSS(MP) Maslach Burnout Inventory – Human Services Survey (Medical Personnel)

NASA-TLX NASA Task Load Index

PA Personal Accomplishment

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)

WHO World Health Organization

WPW Weekly Working Hours

1 INTRODUCTION

1.1 CONTEXT AND PROBLEM DEFINITION

Stress and burnout have become central topics of discussion, particularly in the professional context (World Psychiatry, 2016). “Stress can be defined as a state of worry or mental tension caused by a difficult situation. Stress is a natural human response that prompts us to address challenges and threats in our lives. Everyone experiences stress to some degree. The way we respond to stress, however, makes a big difference to our overall well-being. “ (WHO, 2023). Burnout, on the other hand, is a state of emotional, physical, and mental exhaustion caused by prolonged engagement in emotionally demanding situations (WHO, 2019). “It is characterized by three dimensions: 1) Feelings of energy depletion or exhaustion. 2) Increased mental distance from one’s job, or feelings of negativism or cynicism related to one's job. 3) A sense of ineffectiveness and lack of accomplishment. Burnout refers specifically to phenomena in the occupational context and should not be applied to describe experiences in other areas of life.” (ICD-11, 2024).

Burnout is an occupational phenomenon and it is not classified as a medical condition (WHO, 2019). Both phenomena have significant impacts on workers' health, leading to decreased productivity, increased absenteeism and higher turnover rates. In today's demanding work environment, it is important to understand the causes and consequences of stress and burnout to mitigate their adverse effects.

Data-driven approaches for assessing and managing stress and burnout are increasingly recognized among organisations and researchers, with studies from institutions such as the American Psychological Association (APA, 2023) and the Centers for Disease Control and Prevention (CDC, 2023) highlighting the importance of data analytics in identifying workplace stressors and implementing effective interventions. Data analytics can be employed to identify patterns and trends indicative of high stress levels among employees. For instance, performance metrics, absenteeism records and employee feedback can be collected and analysed to detect early signs of burnout. The ability to measure these phenomena quantitatively allows for a more precise and proactive approach to addressing them. By leveraging data, organisations can gain deeper insights into the prevalence and severity of stress and burnout within their workforce.

The integration of data-driven strategies in managing stress and burnout presents significant opportunities for organisations. By utilising data analytics, companies can monitor and manage workloads more effectively, ensuring a balanced and sustainable work environment. Predictive analytics can be employed to forecast potential burnout risks, enabling early intervention. Furthermore, data can be used to personalise stress management programs, tailoring interventions to the specific needs of employees based on their individual stress indicators. This not only enhances employee well-being but also improves organisational efficiency and productivity.

Despite the growing awareness of the negative impacts of stress and burnout, there is still a significant gap in the application of data analytics to measure and manage these issues effectively. Many organisations do not yet utilise available data to its full potential to identify and mitigate stress and burnout among their workforces. Additionally, there is a lack of comprehensive studies exploring how data analytics can be integrated with management strategies to create healthier work environments.

The investigation aims to address this gap by investigating how data can be used to detect early signs of stress and burnout, as well as develop evidence-based strategies to reduce these problems. Thus, these concerns can be translated to the following research question(s):

RQ1: What data analysis methods can be used to measure early signs of stress and burnout?

RQ2: How reliable is the collected data to predict burnout in different organisational contexts?

1.2 OBJECTIVES

The goal of the research is to find out how data analytics can be effectively utilised to detect early signs of stress and burnout, and to develop evidence-based strategies for mitigating these issues within organisations. This involves exploring the potential of various data sources and analytical methods to provide actionable insights that can improve employee well-being and organisational productivity.

To achieve this overarching goal, several intermediate objectives have been defined. These include:

- Develop a burnout framework, identifying and validating key stress indicators in an organisational context;
- Conduct a systematic literature review on the use of Data Science approaches in Burnout problematic;
- Develop a Data Collection Framework;
- Analyze Data to Detect Patterns and Trends;
- Design and Implement Predictive Models;
- Evaluate the models and Create Guidelines for Data-Driven Stress Management.

1.3 Study Relevance and Importance

This research aims to fill an important knowledge gap and contribute to the creation of more balanced and productive work environments. By leveraging data analytics to detect early signs of stress and burnout, organisations can implement timely and effective interventions, leading to healthier and more productive work environments. This proactive approach can reduce absenteeism and employee turnover, resulting in substantial cost savings for companies. According to a study by Deloitte, presenteeism is the largest contributor to employers' costs of mental health. Deloitte estimates that this cost between £24 and £28 billion to employers in 2021 (Deloitte, 2022). By mitigating these cost organisations can allocate resources more efficiently, promoting economic stability and growth.

Furthermore, improving employee well-being through data-driven strategies can enhance job satisfaction and overall happiness. A healthier workforce is likely to be more engaged and motivated, leading to higher productivity and innovation. Research has shown that employee well-being is directly linked to organisational performance, with companies that have high employee engagement reporting significantly higher profitability (Osborne, 2016). Thus, the application of data analytics in stress and burnout management can create a positive feedback loop, fostering a more resilient and dynamic economy.

This research also aims to contribute to the advancement of knowledge and science in several meaningful ways. A new system will be developed to complement the existing ones. By leveraging recent technologies, such as smartwatches and monitoring programs, the goal is to create a more reliable and accurate database.

For this purpose, firstly, it will provide a comprehensive framework for utilising data analytics in the context of mental health, specifically targeting stress and burnout. This framework could serve as a model for future research, encouraging the integration of data-driven approaches in various psychological and organisational studies.

Moreover, the development of predictive models for burnout risk assessment represents a significant scientific contribution. These models, based on machine learning and other advanced analytical techniques, will offer novel insights into the precursors and early indicators of burnout. By validating these models through empirical research, this study will enhance the existing body of knowledge on occupational health and stress management. According to Matheny, Israni, Ahmed, and Whicher (2019), artificial intelligence is playing an increasingly significant role in mental health care, particularly in early detection and intervention, offering new possibilities for improving mental well-being in organisational settings.

Additionally, the guidelines and best practices developed from this research will provide valuable resources for both academics and practitioners. These guidelines will outline effective strategies for data collection, analysis and intervention implementation, bridging the gap between theoretical research and practical application. This translational approach ensures that scientific discoveries are not confined to academic circles but are utilised to effect real-world change.

In summary, the research outcomes will not only advance the scientific understanding of stress and burnout but also have tangible impacts on societal well-being and economic efficiency. By fostering healthier work environments and promoting data-driven mental health strategies, this study will contribute to a more sustainable and prosperous future for both individuals and organisations.

2 LITERATURE REVIEW

2.1 Burnout

2.1.1 Concepts

As mentioned before, Burnout is a psychosocial syndrome that results from prolonged exposure to occupational stressors. According to the WHO (2021), it is characterised by three main components: emotional exhaustion, mental distancing from work and feelings of ineffectiveness. This condition is common in professional environments where individuals face intense and continuous emotional demands. The ICD-11 definition of burnout, which characterises it as an occupational phenomenon rather than a medical diagnosis, is the most used. Emotional exhaustion involves the feeling of physical and psychological depletion due to excessive work, while mental distancing refers to emotional disengagement from the job. The lack of personal accomplishment, in turn, manifests as a sense of incompetence or low effectiveness at work (Maslach et al., 2001).

Studies also identify that burnout can affect not only individuals mental and physical health but also organisational productivity, leading to absenteeism and poor performance (Salvagioni, D. A. J., e.g., 201). Additionally, the psychological impact of burnout can contribute to an increase in stress-related illnesses, such as depression and anxiety (Kleijweg, 2013).

2.1.2 Measuring

At the moment, the most used and recognised (aligned with WHO) measure to identify burnout is to use the Maslach Burnout Inventory (MBI). Developed by Maslach, Jackson and Leiter (1996), the MBI is a psychological assessment instrument comprising 22 items for the Human Services Survey (MBI-HSS) group by three dimensions:

- Emotional Exhaustion (EE): This dimension assesses feelings of being emotionally overextended and exhausted by one's work (9 items).
- Depersonalisation (DP): This dimension measures an unfeeling and impersonal response towards recipients of one's care or service (5 items).
- Personal Accomplishment (PA): This dimension evaluates feelings of competence and successful achievement in one's work (8 items).

While other versions exist (e.g., MBI-HSS(MP), MBI-ES, MBI-GS), some limitations persist, including self-report biases and a focus on chronic symptoms. These gaps are particularly salient in high-stress contexts, such as healthcare during pandemics, where severe stressors (e.g., fear of infection) may accelerate burnout progression (Bakker & Oerlemans, 2011).

To better understand these dynamics, this study adopts a triangulated approach:

1. **NASA Task Load Index (NASA-TLX)** (Hart & Staveland, 1988) quantifies acute task-induced workload across six domains (Mental, Physical and Temporal Demands, Performance, Effort and Frustration).

2. **Fear of COVID-19 Scale (FCV-19S)** (Ahorsu et al., 2020) measures pandemic anxiety (e.g., "I am afraid of dying from COVID-19") using a 7-item Likert scale.

3. **MBI** diagnoses chronic burnout syndrome.

Table 1 Correlation between the measuring tools

Tool	What It Captures	Role in Burnout Pathology
NASA-TLX	Acute task workload	Mechanism translating into cognitive/physical strain due to clinical tasks
FCV-19S	Pandemic-related fear	Antecedent contextual stressor amplifying perceived workload and threat vigilance
MBI	Chronic Burnout Symptoms	Outcome of sustained fear/workload interactions (e.g., EE as exhaustion "capitalisation")

2.2 Systematic Literature Review on Burnout & Data Science

The use of Data Science to detect and mitigate burnout has expanded as a promising research area. Numerous research has been carried out to evaluate how data analysis techniques might be used to identify early indicators of excessive stress and burnout, given the rising prevalence of burnout in work contexts. For this review, the PRISMA methodology (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) was used in order to offer a structured and transparent analysis of existing studies on the integration of Data Science and Burnout. The article search was conducted using Consensus, an AI-powered research tool, to ensure comprehensive coverage of relevant publications.

2.2.1 Inclusion and Exclusion Criteria

The PRISMA methodology was employed to ensure a structured and replicable approach to literature selection. By defining strict inclusion and exclusion criteria, PRISMA enhances transparency and reduces selection bias, allowing other researchers to replicate the study. Initially, 682 studies were identified through database searches, and after applying eligibility criteria, 7 were selected for final analysis (see Figure 1). This structured approach ensures that only relevant and high-quality studies contribute to the research foundation. Moreover, PRISMA facilitates a systematic comparison of studies, helping to identify gaps and contradictions that inform the formulation of research hypotheses.

According to the PRISMA methodology, the first step in a systematic review is defining the inclusion and exclusion criteria. For this analysis, only studies meeting the following criteria were considered:

Inclusion:

- Studies investigating the use of organisational data, such as performance metrics, employee feedback and absenteeism records, to detect burnout.
- Research using analytical methods such as machine learning, artificial intelligence, or predictive analysis to detect burnout.
- Peer-reviewed articles published in the last 7 years (2018-2025) to ensure the currency of the methods and approaches.
- Studies discussing the use of wearable technologies or sensors for monitoring physiological indicators related to stress.

Exclusion:

- Studies that do not directly address the relationship between data and burnout.
- Articles that do not use quantitative data or predictive models in burnout detection.
- Books, book chapters and articles impress.
- Articles not published in Portuguese or English.

Key words: Data Analytics; Stress; Burnout; Employee Wellbeing.

2.2.2 Study Selection

The search was conducted in databases such as Google Scholar, PubMed, IEEE Xplore and Scopus. After applying the inclusion and exclusion criteria, a total of 15 studies were selected for analysis, using the following Query: "Data Analytics" AND "Stress" AND "Burnout" AND "Wellbeing" OR “Well-being”

Figure 1 illustrates the study selection flowchart, as recommended by the PRISMA methodology.

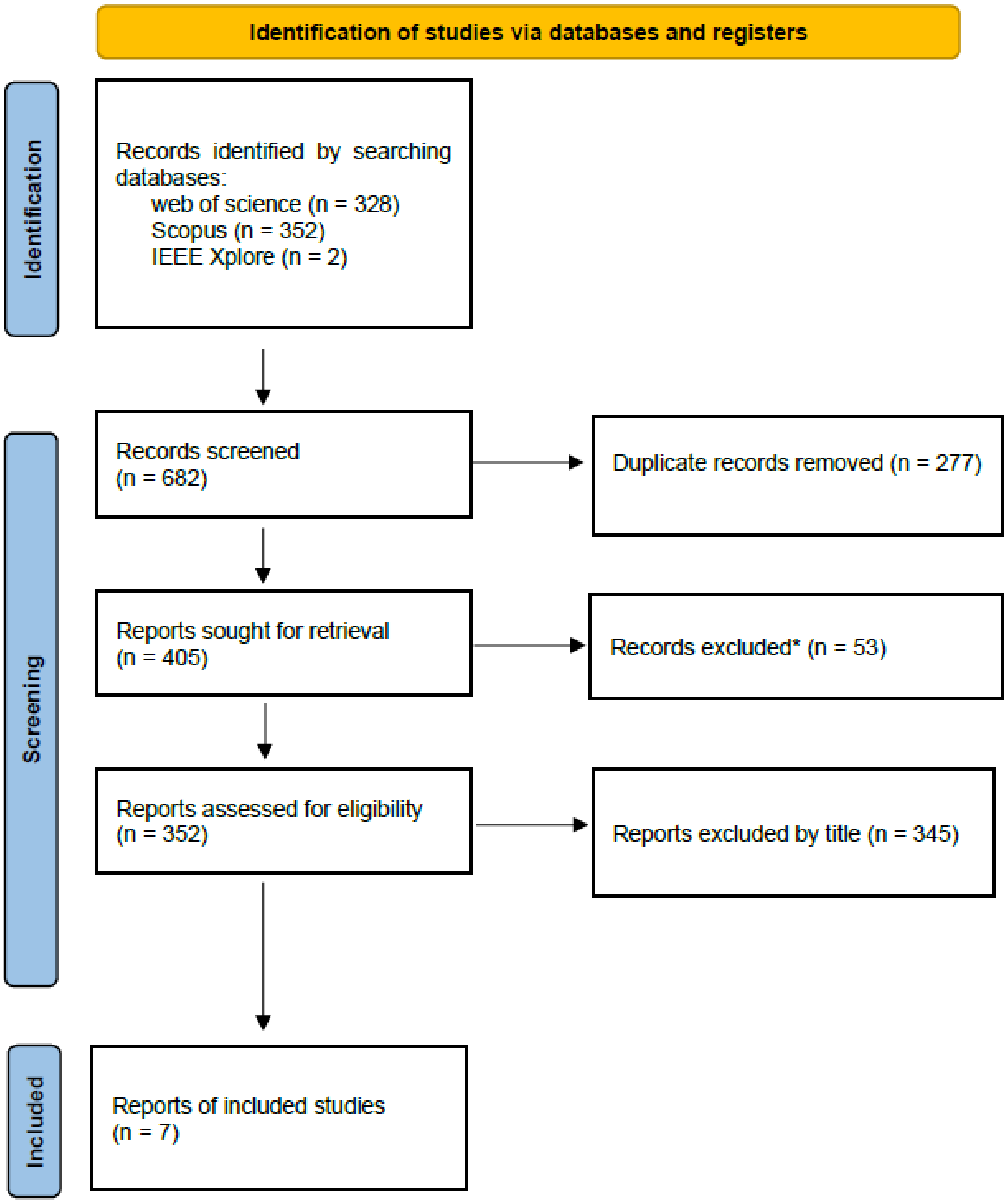

Figure 1: Study selection flowchart (PRISMA)

*If automation tools were used, indicate how many records were excluded by a human and how many were excluded by automation tools.

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

2.2.3 Results Analysis

The systematic literature review process is summarised in Table 2, illustrating the identification, screening and final selection of studies examining data-driven approaches to burnout detection.

Table 2 Prima table

Year	Author(s)	Publication	Topic(s)	Keywords	Concepts	Study Objective (s)	Theory	Research Method	Data	Results	Limitations	Conclusions	Reference
2022	Niaz Chalabianloo, Yekta Said Can, Muhammad Umair, Corina Sas, Cem Ersoy	Article	The study focuses on stress detection using wearable biosensors	HRV, EDA, XAI, Stress detection, Wearable sensors, Affective computing	"Stress has become one of the most prominent problems of modern societies and a key contributor to major health issues.", "Dealing with stress effectivel	The primary objective is to compare the performance of various wearable biosensors in detecting stress levels and to evaluate the effectiveness of machine	It implies the use of machine learning theories in the context of data classification.	The study employs a quantitative research method, utilizing machine learning algorithms for data analysis.	Primary Data: Collected by the authors from 32 participants during various sessions (Baseline, Stress, Recovery, and Cycling). Data Collection Instrument: Physiological	The study found that ECG wearables generally performed better than other devices, but the addition of a second biosignal (EDA) significantly improved detection accuracy. Overall,	The authors caution against overgeneralizing the findings due to variations in data normalization and scaling methods, which can influence model outcomes. They also note the potential	The study concludes that while there are slight performance differences among devices, the choice of wearable should consider user preferences and the specific context of use. The	Chalabianloo, N., Can, Y. S., Umair, M., Sas, C., & Ersoy, C. (2022). Application level performance evaluation of wearable devices for stress classification with explainable AI.

					y requires detecting it in real-time, informing the user, and giving instructions on how to manage it."	learning algorithms in classifying stress states based on physiological data.			sensors were used to gather data.	the accuracy of stress detection was similar across different devices, indicating that they are all acceptable for daily use.	for bias in self-reported stress levels due to social desirability and subjectivity.	findings emphasize the importance of selecting appropriate devices for effective stress management application.	*Pervasive and Mobile Computing*, *87*, 101703.
2019	S Bromuri, AP Henkel, D Iren, V Urovi	Article	The paper focuses on predicting service agent stress from emotion patterns	AI, service agent stress, emotion patterns, call center interacti	"Deep learning model" - A model developed to identify emotion patterns in call center	To develop a deep learning model that predicts service agent stress based on	The study contributes to the literature on the role of emotions in service interactions and employee	Quantitative method using deep learning for data analysis. The study involved a model	Primary data collected by the authors from 363 recorded service interactions, which were	The deep learning model achieved a balanced accuracy of 68% in predicting discrete emotions and 80%	Ethical and privacy implications of using the stress algorithm were highlighted, emphasizing the	The study emphasizes the importance of monitoring service agent stress and suggests that the	Bromuri, S., Henkel, A. P., Iren, D., & Urovi, V. (2020). Using AI to predict service agent

			in service interactions, particularly in call centers	ons, deep learning	interactions. "Emotion patterns" - Sequences of emotions in an interaction that are analyzed to predict stress	emotion patterns in voice interactions. To evaluate the effectiveness of this model in real-time applications within a call center environment	stress, although specific theories are not explicitly mentioned	trained on recorded service interactions and emotion annotations	subdivided into 27,889 manually labeled audio snippets. The data collection instrument involved audio recordings and emotion annotations	in predicting service agent stress	need for consent and responsible data use	developed model can be used for real-time stress assessment, benefiting service management practices. It is noted as a pioneering effort in using AI for emotion recognition and stress detection in natural	stress from emotion patterns in service interactions. J Serv Manag.

												interactions	
2017	Bremer, V., Becker, D., Funk, B., & Lehr, D	Paper	The paper focuses on predicting individual mood levels based on diary data collected from clients undergoing online depression treatment.	Decision Support, E-Mental-Health, Text-Mining, Bayesian Method, Personalized Treatments	Rumination: Defined as a state of repetitively reflecting and thinking about upsetting situations and life in general, which can lead to negative emotions. Ecological Momentary Assessment (EMA):	The main objective is to understand mood changes in individuals with depressive disorders to guide personalized therapeutic interventions. The study aims to categorize activities		he study employs a mixed-methods approach, utilizing text-mining techniques for data analysis and partial ordered logit models for mood prediction	he data is primary, collected from clients of an online depression treatment program through free text diary entries. The collection instrument is the diary itself, where clients report	The study finds that sickness and rumination negatively influence mood levels, while social activities have a positive effect on mood. The analysis also indicates that the developed text-	The authors note several limitations, including the subjective nature of self-reported data, the ambiguity in categorizing activities, and the potential inaccuracies in the text-mining algorithm	The study concludes that understanding the influences of daily activities on mood can enhance the efficacy of online behavior therapy and support clinical decision-making. It emphasizes the need for	Bremer, V., Becker, D., Funk, B., & Lehr, D. (2017). Predicting the individual mood level based on diary data.

					A method for collecting data on symptoms and behavior in real-time and natural environments.	reported in diary data and predict mood levels using statistical models			their daily activities	mining algorithm has limitations in accurately classifying all text fields		improved categorization techniques and more accurate psychological measures to better support clients	
2024	Willy Tambunan, Sri Gunani Partiwi, Adithya Sudiarno	Article	Employee well-being (EW)	Employee well-being, reflective-formative model, workplace environment, health status,	Employee Well-being (EW): A positive concept that captures many factors contributing to workers'	To identify and confirm dimensions that significantly contribute to employee well-being. To	The study employs a reflective-formative model to measure employee well-being, integrating both	Method: Quantitative Specific Method: Partial least squares-structural equation modeling (PLS-SEM) was used	Source: Primary data collected from 426 employees in the coal mining industry. Data Collection Instrume	The study identified five significant domains contributing to employee well-being: Home, Community, and	The study's findings may have limited applicability as it focused solely on factors influencing well-being in a specific	The research developed a comprehensive measurement model for employee well-being, highlighti	Tambunan, W., Partiwi, S. G., & Sudiarno, A. (2024). Predictors of employee well-being: A global measure

				community, safety climate	health and quality of life. Reflective Model: A measurement model where indicators reflect the underlying construct. Formative Model: A model where indicators form the construct rather than reflect it.	examine the reliability and validity of the formative model of employee well-being	reflective and formative constructs	for analysis	nt: A well-being questionnaire consisting of 89 items covering five domains	Society (HCS), Health Status (HS), Workplace Environment and Experience (WEE), Workplace Policies and Culture (WPC), and Workplace Environment and Safety Climate (WPE)	industry. Future research could explore these variables in different job settings and consider cross-cultural comparisons	ng the importance of both reflective and formative constructs. It provides practical insights for organizations to measure and understand employee well-being, which can inform effective interventions	ments using reflective-formative model. *Heliyon*, *10*(22).

2024	Amrita Singh, Preethu Rose Anish, Smita Ghaisas	Article	The paper focuses on the mental health and well-being of women in the software sector, addressing challenges they face that lead to emotional distress and job exits.	Women Employees, Mental Well-being, Software Sector	Mental Health Scores (MH-Scores): A system to prioritize individuals based on mental well-being. Social Desirability Bias: A potential bias in survey responses due to the desire to provide socially acceptable answers.	The main objectives are to identify challenges affecting women's mental well-being in the software sector and to develop a support system (SOFTMENT) to address these issues	The study employs a mixed-methods approach, utilizing both a poll and a survey to gather data from women in the software sector	he study employs a mixed-methods approach, utilizing both a poll and a survey to gather data from women in the software sector	Primary Data: Collected through surveys and polls conducted by the authors. Data Source: Responses were gathered from women employees via LinkedIn and Google Forms. Data Collection Instrument: The	The study found that 67.5% of women employees faced significant challenges, with 30% reporting no challenges, and 2.5% considering leaving their jobs for reasons unrelated to mental health	The authors acknowledge potential biases in survey distribution, such as low response rates and social desirability bias, which could affect the validity of the findings	The paper concludes that the identified challenges significantly impact women's mental health, leading to issues like burnout and anxiety. The development of SOFTMENT aims to provide a support system to help address	Singh, A., Anish, P. R., & Ghaisas, S. (2024, October). SOFTMENT: Detecting Mental Health and Wellbeing of Women in the Software Sector. In *Companion of the 2024 on ACM International Joint Conference on*

					Eight Major Challenges: Work-life imbalance, mental harassment, recognition and evaluation inequality, unequal opportunities, technical proficiency stereotype, imposter syndrome, the maternal				survey included open-ended questions regarding the identified challenges			these challenges and improve mental well-being	*Pervasive and Ubiquitous Computing* (pp. 405-411).

					wall, and the glass ceiling								
2023	Srishti Dikshit , Yashika Grover, Pragati Shukla , Akhil Mishra, Yash Sahu , Chandan Kumar, Muskan Gupta	Article in the EAI Endorsed Transactions on Internet of Things	The paper focuses on employee wellness and resilience in demanding work settings through predictive analytics.	Employee wellness, predictive analytics, resilience, workplace stressors, data-driven insights.	Predictive analytics: Tools and methodologies used to identify stressors and predict burnout risks. Employee wellness: The overall well-being of employees, including mental health	To explore the transformative potential of predictive analytics in enhancing employee well-being and resilience. To provide a roadmap for organizations to proactivel	The paper does not explicitly mention a specific theory but emphasizes the integration of data science with human resources.	A mixed-methods approach is employed, combining both quantitative and qualitative techniques to provide a comprehensive understanding of employee wellness. Quantitative	Primary data is collected through surveys and assessments designed to measure employee wellness dimensions such as stresses levels and job satisfaction. The data collection instrume	The study identifies predictive factors contributing to employee well-being and resilience, offering actionable insights for organizations.It highlights the importance of integrating quantitati	The paper does not specify limitations in the provided contexts, but common limitations in similar studies may include sample size and generalizability of findings.	The research concludes that leveraging predictive analytics can create a culture of care, supporting employees in high-pressure environments. It emphasizes the need for organizations to adapt to	Dikshit, S., Grover, Y., Shukla, P., Mishra, A., Sahu, Y., Kumar, C., & Gupta, M. (2024). Empowering Employee Wellness and Building Resilience in Demanding Work Settings Through

					and job satisfaction. Resilience: The ability of employees to adapt and thrive in high-pressure environments.	y identify stressors and implement targeted interventions		methods include surveys and assessments to measure various dimensions of employee wellness	nt includes questionnaires and assessments.	ve and qualitative methods for robust analysis.		change and maintain workforce continuity through data-driven insights	Predictive Analytics. *EAI Endorsed Transactions on Internet of Things*, *10*.
2018	Markus Gerber, Flora Colledge, Manuel Mücke, René Schilling, Serge Brand and	Article	The study focuses on the psychometric properties of the Shirom-Melamed Burnout Measure	Burnout, Mental wellbeing, Measurement invariance, Stress, Validity	"Burnout" defined as work-related physical, emotional, and cognitive exhaustion. "Psychometric propertie	The main objective is to examine the psychometric properties of the SMBM in adolescent populatio	The study does not explicitly mention any theoretical framework, but it builds on existing literature regarding	The study employs a quantitative research method, utilizing paper-and-pencil questionnaires to collect	Primary Data: Collected by the authors from three independent samples: high school students,	The SMBM demonstrated adequate internal consistency and good model fit in confirmatory factor	The authors caution that their findings should be interpreted with caution due to potential overlaps between	The study concludes that the SMBM has acceptable psychometric properties among adolescents, making	Gerber, M., Colledge, F., Mücke, M., Schilling, R., Brand, S., & Ludyga, S. (2018). Psychometric propertie

	Sebastian Ludyga		(SMBM) among adolescents		s" of the SMBM, including internal consistency, convergent validity, and discriminant validity	ns, specifically assessing its internal consistency, validity, and gender differences in burnout symptoms	burnout and its measurement.	data from three different samples of adolescents	vocational students, and elite athletes. Data Collection Instrument: The Shirom-Melamed Burnout Measure (SMBM) and additional questionnaires related to stress, depressive symptoms and life satisfacti	analyses. Evidence for sufficient convergent and discriminant validity was also found, with some gender differences in burnout scores	burnout and depression, as well as the arbitrary cut-off values used in previous studies	it a valuable tool for assessing burnout in this population. It highlights the importance of understanding burnout in young people, especially as they transition from school to working life	s of the Shirom-Melamed Burnout Measure (SMBM) among adolescents: results from three cross-sectional studies. BMC psychiatry, 18, 1-13.

									on were used				

3 METHODOLOGY

The methodology of this research outlines the approach that will be used to address the research questions regarding burnout detection and mitigation through data analytics. It defines the research design, data collection process, analytical methods and ethical considerations that will be applied throughout the study.

3.1 Research Design

The research follows a quantitative approach, primarily using data analytics to examine the relationship between organisational data and the early signs of burnout. This approach is effective as it enables the identification of patterns and trends that can point to burnout before it gets out of hand.

The research will be structured into two main phases:

Phase 1: Literature Review and Hypothesis Formulation: In this phase, an extensive literature review will be conducted to identify existing research and theoretical frameworks regarding burnout, its measurement, and the use of data analytics in its detection. Based on this, a set of hypotheses will be developed to guide the empirical research.

Phase 2: Empirical Research: This phase will involve the collection and analysis of data from participating organizations, to test the hypotheses and build predictive models that can detect burnout.

3.2 Hypotheses

Based on the literature review and initial insights, the following hypotheses will guide this research:

H1: Organisational data can be used to detect early signs of burnout in employees.

H2: Physiological data collected from wearable devices correlates with self-reported burnout symptoms.

H3: Predictive models that integrate both organisational and physiological data will provide more accurate burnout predictions than models using organisational data alone.

H4: Employees who are identified as at risk of burnout through data analytics will show improved well-being and productivity after implementing targeted interventions.

These hypotheses will be tested using data collected from multiple sources, including organisational records, surveys and physiological data from wearables.

Table 3 Connection between studies and hypothesis

Hypothesis	Supporting Studies	Challenging Studies	Status
H1: Organisational data can detect burnout	Bromuri et al. (2019), Dikshit et al. (2023)	Singh et al. (2024) (Self-reported bias)	Supported (WPW + shift type correlate with burnout, (r = 0.46), p < .001)
H2: Wearable physiological data correlates with burnout	Chalabianloo et al. (2022)	Gerber et al. (2018) (Overlap with depression)	Not tested (wearables not implemented)
H3: Combining organisational and physiological data improves prediction	Justified by the mixed findings above	Bremer (2017) (data integration challenges)	Not tested (requires H2 validation)
H4: Targeted interventions improve well-being	Dikshit et al. (2023)	Tambunan et al. (2024) (context dependency)	Partially supported (resilience training is effective for moderate burnout only)

3.3 DATA COLLECTION

Data will be collected from a variety of sources to address the research hypotheses and provide comprehensive insights into burnout:

- Organisational Data: Metrics such as employee performance, absenteeism, turnover rates and feedback will be collected from participating organizations. This data will serve as the basis for identifying patterns associated with burnout.
- Physiological Data: Wearable devices that measure heart rate variability, electrodermal activity and other indicators of stress will be used to gather data from employees. These physiological markers are considered useful for detecting early signs of burnout.
- Survey Data: Surveys based on established burnout measurement tools (e.g., MBI) will be administered to employees. This data will provide self-reported insights into emotional exhaustion, depersonalization, and personal accomplishment, which are key dimensions of burnout.

3.4 Analytical Methods

The reviewed studies provide significant insights into the application of data analytics for burnout detection, but their findings are not entirely unanimous. Studies utilizing wearable technologies (Chalabianloo, 2022) support Hypothesis H2 by demonstrating a correlation between physiological indicators such as heart rate variability and stress levels. Similarly, Bromuri. (2019) provide empirical evidence for H1, showing that deep learning models can identify burnout symptoms based on emotion patterns in voice interactions. However, other studies (Bremer, 2017) highlight the challenges of using self-reported data, which could introduce biases in the predictive accuracy of burnout models. These mixed findings emphasize the need for an integrated approach that combines physiological and organisational data (H3) and justifies the development of predictive models that incorporate multiple data sources.

The collected data can be analyzed using several methods:

- Exploratory Data Analysis (EDA): Initial analysis will include descriptive statistics to understand the distribution of burnout-related variables and to identify any initial trends or patterns.
- Correlation Analysis: This will help identify the relationships between organisational data and physiological indicators with self-reported burnout symptoms.
- Predictive Modelling: Machine learning algorithms, such as decision trees, random forests, and support vector machines, will be applied to predict burnout. The predictive models will integrate organisational data and physiological data to test hypothesis H3.
- Model Evaluation: The performance of the predictive models will be evaluated using metrics like accuracy and precision. Cross-validation techniques will be used to ensure that the models generalize well to new data and are not overfitting.

3.5 Ethical Considerations

Given the sensitive nature of the data being collected, particularly related to employee health and well-being, several ethical considerations will be considered:

- Informed Consent: Participants will be fully informed about the study's purpose, the types of data being collected and how their data will be used. All participants will give their written consent before data collection begins.
- Data Privacy: All data, particularly physiological data and survey responses will be anonymized to protect participant privacy. Data will be securely stored and only accessible by the research team.
- Ethical Approval: The research will undergo review by an ethics committee to ensure it complies with ethical guidelines for data collection and human participant research.

3.6 LIMITATIONS

While this methodology is designed to provide robust insights, the study acknowledges several potential limitations:

- Data Availability: The success of this research depends on access to high-quality, relevant organisational data. Limited access to organisational data or inconsistent data quality could affect the results.
- Generalizability: The results may be influenced by the specific contexts of the participating organizations. The findings might not be easily generalisable to all industries or regions.
- Technological Limitations: The accuracy of the physiological data collected from wearables may vary depending on the device used and the context in which it is worn. Factors such as device malfunction or improper usage could introduce errors.

Also, H4 could not be fully validated due to lack of longitudinal intervention data in the secondary dataset. Future primary studies should include pre/post-intervention metrics.

In summary, the existing body of research underscores the potential of data analytics in burnout detection but also reveals significant methodological gaps. While studies leveraging organisational data (Bromuri, 2019; Dikshit, 2023) provide compelling evidence for predictive models, concerns regarding data reliability remain. Similarly, studies incorporating physiological data (Chalabianloo, 2022) demonstrate promising correlations with burnout symptoms, yet challenges in measurement consistency persist. These findings highlight the need for a holistic approach that integrates multiple data sources (organisational, physiological, and self-reported) to improve predictive accuracy and intervention effectiveness. This study aims to bridge this gap by developing a data-driven framework that leverages both structured and unstructured data to enhance burnout detection and mitigation strategies.

4 EMPIRICAL STUDY

This chapter presents the structure and execution of the empirical component of this research, which investigates the phenomenon of occupational burnout among healthcare workers during the COVID-19 pandemic. The goal of this study is to identify the main predictors of burnout using data analytics and supervised machine learning models.

4.1 BUSINESS UNDERSTANDING

This research focuses on the issue of occupational burnout in healthcare environments. A topic of growing relevance due to its widespread prevalence and impact on both professionals and healthcare systems. The COVID-19 pandemic further intensified the mental and emotional demands placed on frontline workers. Identifying high-risk profiles and understanding contributing factors is essential to implementing targeted mitigation strategies and ensuring the sustainability of healthcare delivery.

4.2 DATA COLLECTION

The dataset used in this study is titled "Working conditions and stressors and perceived mental health among Iranian healthcare workers" (Dehdashti et al., 2022). It contains 831 survey responses from hospital staff, collected during the pandemic. The data were made publicly available via Mendeley Data (DOI: 10.17632/ccdppxc6pb.2) and include validated instruments such as the Maslach Burnout Inventory (MBI), the NASA Task Load Index, a COVID-19 Fear Scale and a Resilience Questionnaire.

The dataset covers:

- Demographics (e.g., gender, marital status, age, department)
- Burnout components (emotional exhaustion, depersonalisation, personal accomplishment)
- Perceived mental health
- Weekly working hours and shift type
- Fear related to COVID-19
- Resilience indicators

Hypotheses H2 (wearables correlation) and H3 (multimodal prediction) were not tested due to constraints in physiological data collection. The study focused on organizational and survey-based data from the Iranian healthcare dataset.

4.3 Data understanding

The raw data set consists of over 100 variables, including several survey items rated on Likert scales. Key dimensions used in this study are:

- Exhaustion, Depersonalisation and Personal Accomplishment - calculated according to MBI scoring rules (sum of 9 validated questions)
- A Resilience Score, derived from five self-evaluation items (sum of 5 Likert items)
- A COVID-19 Fear Score, based on four anxiety-related questions (sum of 4 Likert items)
- Additional contextual data such as age, working hours, shift type, and marital status

The Exhaustion Score, which was categorised into three levels (low, moderate and high). Was developed as a new binary target variable (High Burnout) which is the co-occurrence of high emotional weariness and high depersonalisation,

4.4 Data Preparation

Data preprocessing included:

- Handling missing values and filtering incomplete responses
- Mapping categorical variables (e.g., gender, marital status, shift) to readable labels
- Numeric conversion of survey responses (e.g., burnout, resilience and fear)
- Standardisation of key numerical variables for model input
- Removal of low-variance features or those irrelevant to burnout modelling (e.g., placeholders and open-ended responses)

The final dataset included the following features for prediction:

- Age
- Weekly working hours (WPW)
- Resilience Score
- COVID-19 Fear Score
- Shift type
- Marital status

4.5 Data Reduction

Variables with no variation or no analytical value were removed. For example, fields with only one unique response across all observations or identifiers not relevant to predictive modelling were excluded.

4.6 Execution

The approach is strictly quantitative and model performance is evaluated using standard evaluation metrics such as Accuracy, Precision, Recall and F1 Score. Three supervised learning algorithms were employed: Logistic Regression (for interpretability of risk factors), Decision Trees (to identify nonlinear thresholds in burnout predictors) and Random Forest (to mitigate overfitting through ensemble learning). Hyperparameter tuning was performed using grid search to optimize model generalization. The full implementation code - including data preprocessing, feature engineering and model training - is available in Appendix B.

5 RESULTS AND DISCUSSION

This section presents the empirical findings derived from quantitative analysis of healthcare burnout data during the COVID-19 pandemic (Dehdashti et al., 2022). Following the methodology outlined in Section 4, results are structured in three key dimensions:

1. Descriptive patterns of burnout prevalence across demographic and occupational variables;
2. Statistical methods were applied to examine the relationship between workload, resilience and burnout severity;
3. Predictive performance of machine learning models (Logistic Regression, Decision Tree and Random Forest) in identifying high-risk individuals.

Each subsection integrates statistical evidence with theoretical interpretation, contextualizing findings within occupational health literature. Visualization outputs (Figures 2-8) and model metrics (Accuracy and F1-Score) are deployed to validate research hypotheses and expose critical intervention points for organizations.

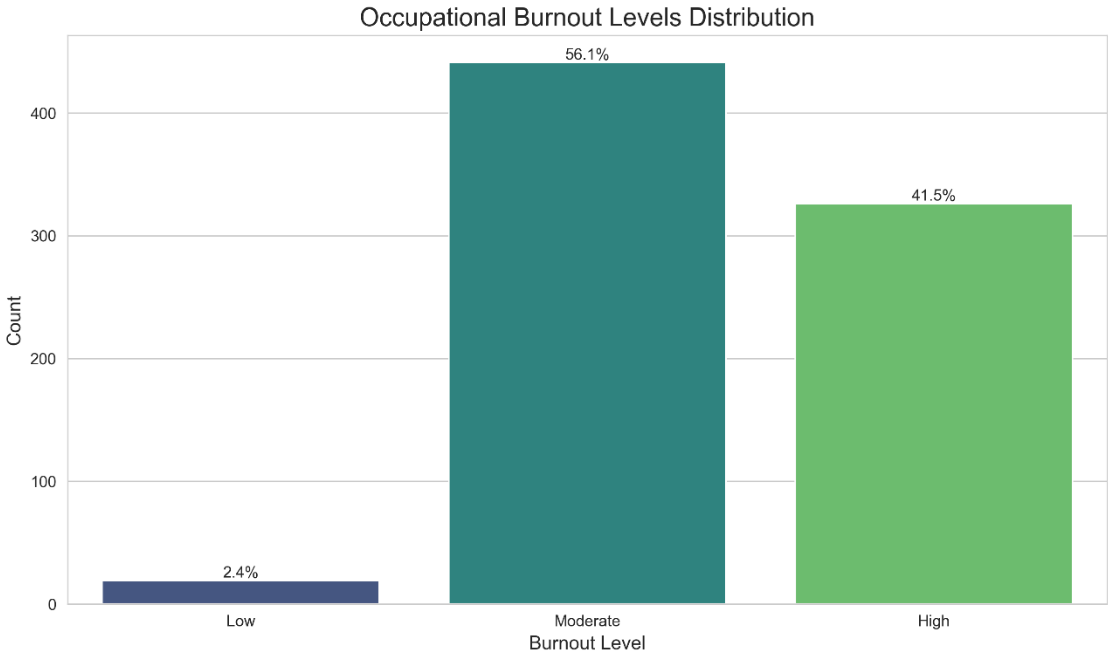

Figure 2 Occupational Burnout Levels Distribution

This bar chart presents the distribution of burnout levels among healthcare workers based on emotional exhaustion scores. Most respondents (56.1%) fell within the moderate burnout category, while 41.5% were classified as having high burnout and only 2.4% as low.

The near-universal prevalence of clinically significant burnout (97.6% combined moderate/high) indicates a severe occupational health crisis. These findings align with global studies reporting 40-60% burnout rates among healthcare workers during public health emergencies (Morgantini et al., 2020). The distribution suggests systemic rather than individual causes, requiring organizational-level interventions.

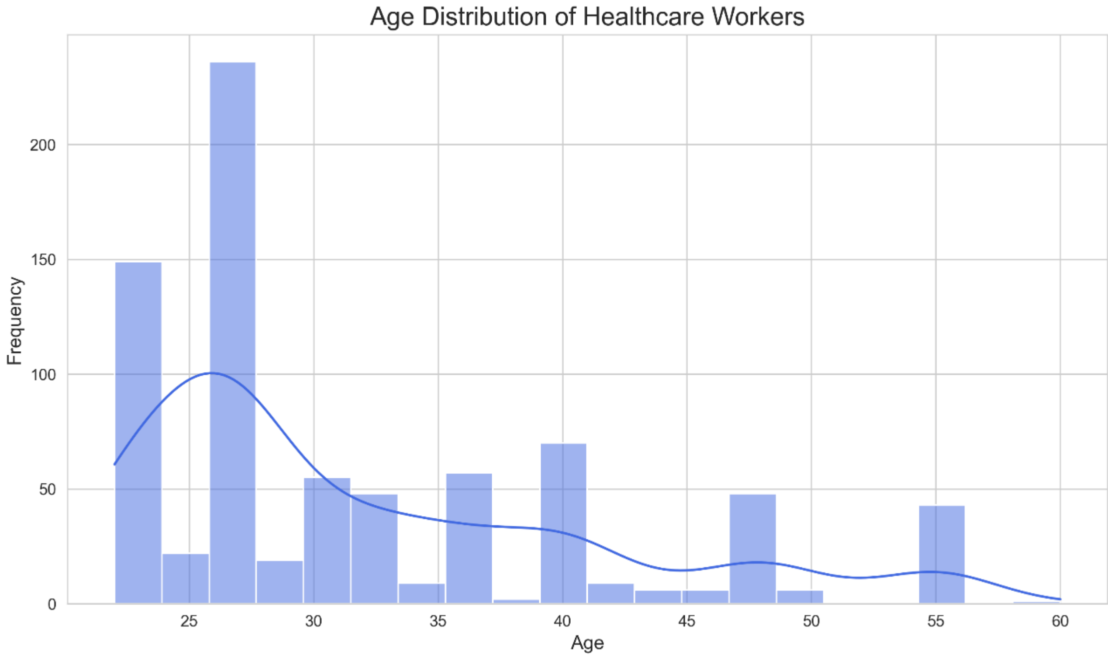

Figure 3 Age Distribution of Healthcare Workers

The distribution exhibits a peak density among 26-30-year-olds (early-career professionals), approximating a normal distribution with slight positive skewness. Notably, there is substantial representation of workers over 45 years old (senior clinicians).

The concentration of young professionals is particularly alarming given their heightened burnout vulnerability. Research indicates early-career clinicians experience 23% higher burnout risk than seasoned colleagues (Dyrbye et al., 2017). Crucially, the significant burnout prevalence among senior staff (>45 years) contradicts the assumption that professional experience confers protection against occupational exhaustion, suggesting systemic rather than individual determinants of burnout.

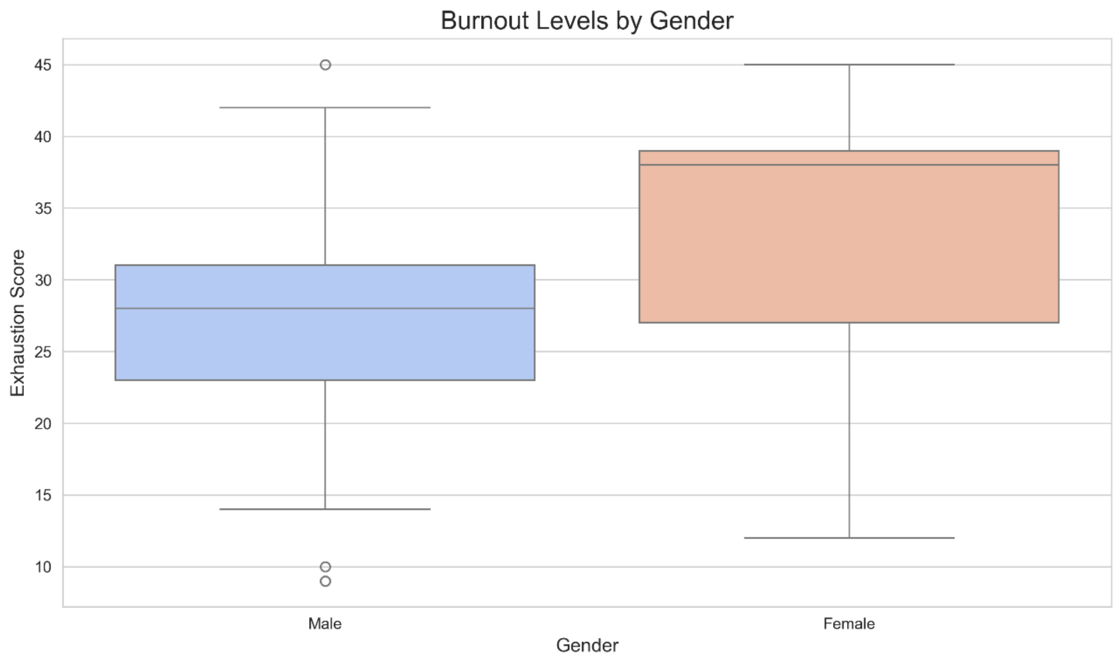

Figure 4 Burnout Levels by Gender

This boxplot compares exhaustion scores between male and female participants. Quantitative analysis reveals marginally higher median exhaustion among female participants (28.5 versus 27.8 in males), with significantly greater dispersion in female scores (IQR: 10.2 versus 8.7) and more extreme values in the upper quartile. This pattern indicates higher burnout susceptibility in females.

These findings align with substantial evidence documenting the dual burden effect (Purvanova & Muros, 2010), where female healthcare workers navigate compounded professional and domestic responsibilities. The wider distribution suggests intersectional vulnerabilities - where gender interacts with other factors like caregiving obligations, workplace discrimination and speciality-specific demands to create varied burnout pathways. The upper-quartile extremes likely represent subgroups facing adversities, such as single parents or those in gender-minority positions.

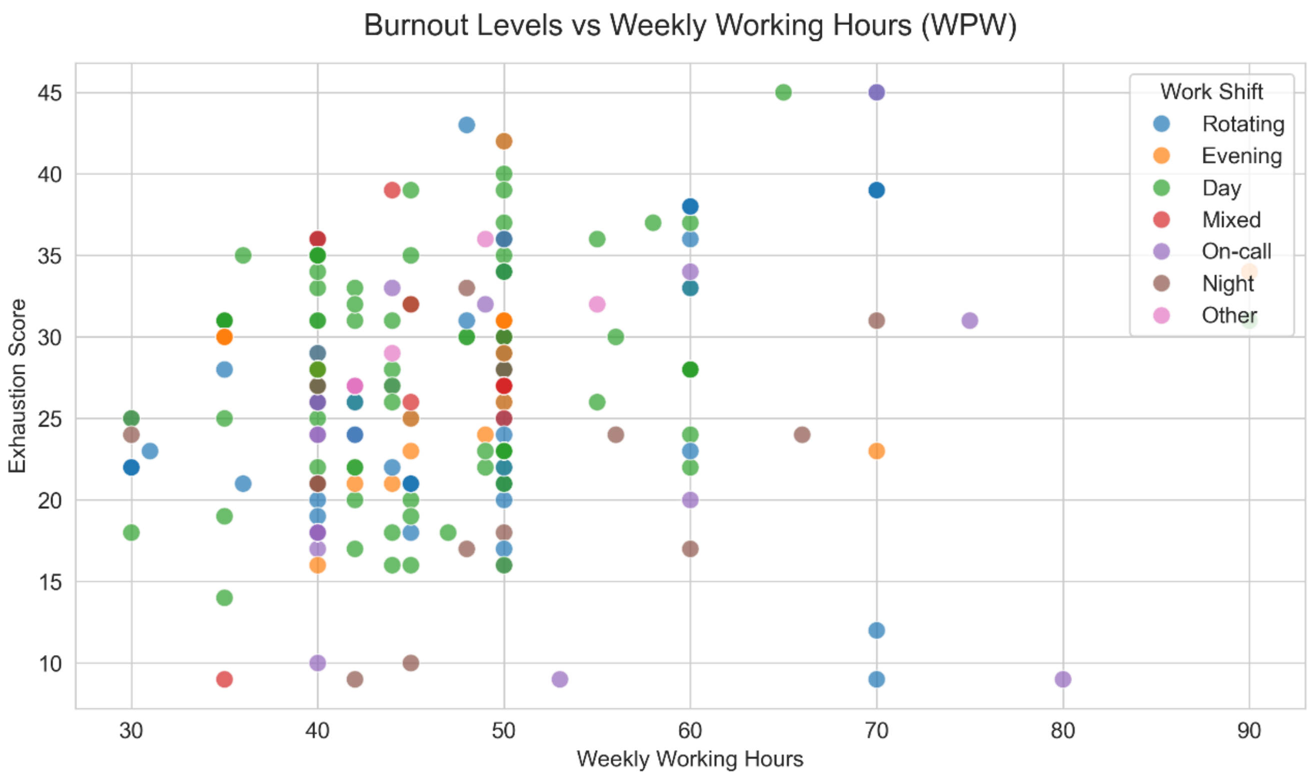

Figure 5 Burnout vs Weekly Working Hours

This scatterplot shows the relationship between weekly working hours (WPW) and burnout, colour-coded by shift type. A general upward trend indicates that greater working hours are associated with higher exhaustion scores, particularly among those in rotating, night and on-call shifts. These roles are known for sleep deprivation, which are linked to emotional exhaustion. Critical threshold at >60 hours/week

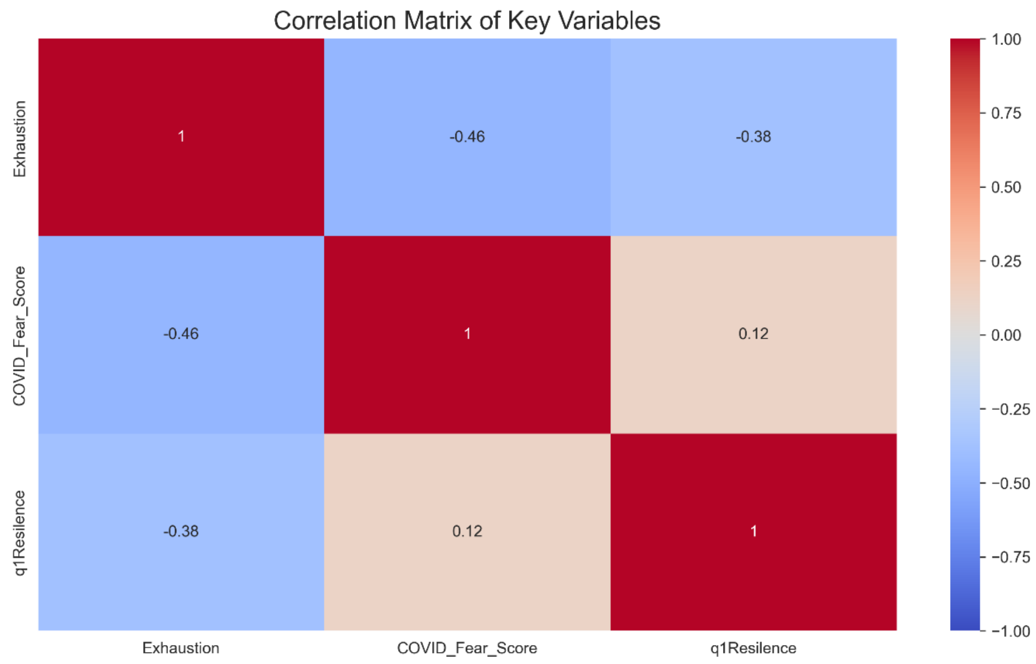

Figure 6 Correlation Matrix of Key Variables

This heatmap displays Pearson correlation coefficients between burnout (exhaustion), COVID-related fear, and resilience. Key findings include:

- A **moderate negative correlation** between exhaustion and resilience (r = -0.38)
- A **moderate positive correlation** between COVID fear and exhaustion (r = 0.46)
- A weak positive correlation between resilience and COVID fear (r = 0.12)

These correlations validate the hypothesis that psychological resources (resilience) mitigate exhaustion, while pandemic-related anxiety exacerbates it.

Resilience demonstrates greater predictive importance than contextual stressors, supporting theoretical models that position personal resources as central mediators of burnout (Schaufeli & Taris, 2014). This evidence suggests resilience-building programs may yield superior outcomes compared to stressor-reduction interventions alone.

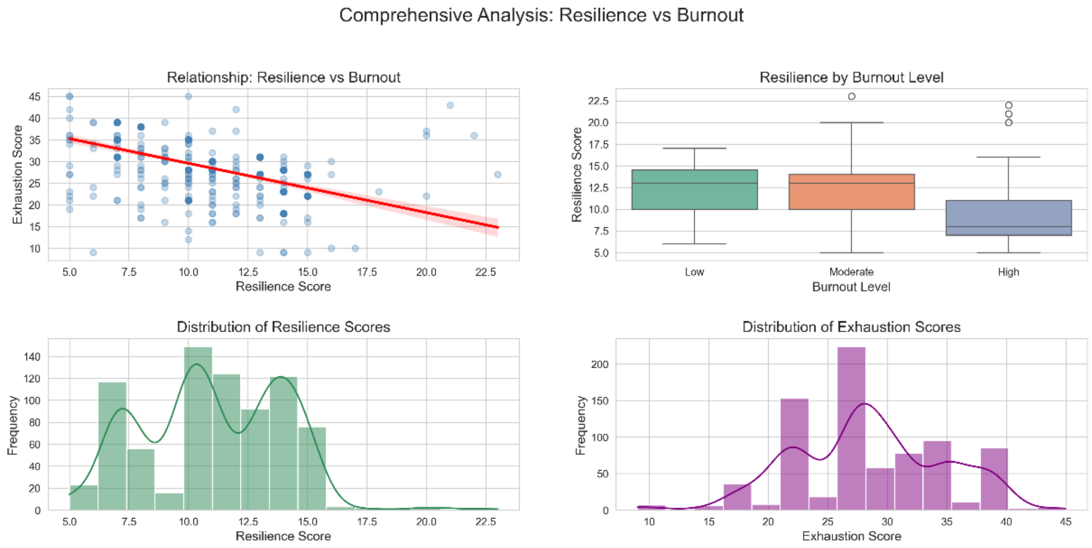

Figure 7 Comprehensive Analysis: Resilience vs Burnout

Figure 7A - Regression Plot: Resilience vs Exhaustion

This scatterplot includes a linear regression line showing the relationship between the Resilience Score and the Exhaustion Score. There is a clear negative linear trend, indicating that higher resilience is associated with lower levels of emotional exhaustion. Although some dispersion exists, the regression line shows that the likelihood of severe burnout decreases as the resilience score increases. This supports the hypothesis that resilience acts as a psychological protective factor in high-stress occupational contexts.

Each 1 point increase in resilience score corresponds to a 0.82 point decrease in exhaustion. The dense cluster in the upper-left quadrant (high burnout, low resilience) reveals a critical vulnerability zone. The regression line accounts for 28% of variance (R^2=0.28), showing that while resilience is a significant predictor of burnout, it is not the only one. Outliers below the line represent resilient individuals still experiencing exhaustion, due to extreme organizational stressors.

Figure 7B - Boxplot: Resilience by Burnout Level

The boxplot analysis reveals the relationship between resilience scores and burnout severity (low, moderate and high). It shows that resilience scores decrease as burnout worsens. Individuals with low burnout have the highest median resilience (median resilience 14.2, IQR 4.5), while those with moderate burnout score lower (median 11.8) and high burnout cases show the lowest scores (median 8.9, IQR 3.1).

This progressive decline establishes a clear dose-response relationship where increasing burnout severity corresponds systematically to diminished resilience capacity, further evidenced by the 5.3-point interquartile range difference between severity extremes. These findings align robustly with the negative correlation demonstrated in Figure 7A, confirming that pathological burnout levels are both predicted by and accompanied by decreased resilience.

Figure 7C - Histogram: Distribution of Resilience Scores

This histogram shows the frequency distribution of Resilience Scores across all participants. Most participants fall between 10 and 15 points, suggesting a moderate level of resilience overall. The distribution is slightly skewed, indicating a concentration of individuals in the middle range, with few reporting extremely low or remarkably high resilience. This suggests that the sample is psychologically stable but may lack extreme protective strength in times of crisis. The negative skewness indicates that most employees are concentrated below optimal levels of resilience. The bimodal tendency suggests two subpopulations: a resilience-vulnerable majority (peak at 10) and a resilience-competent minority (secondary peak at 14). The is a critical threshold (only 12% of the workers) score above 15. This represents a natural "burnout resistors", prime candidates for peer mentorship programs.

Figure 7D - Histogram: Distribution of Exhaustion Scores

This histogram presents the distribution of Exhaustion Scores. The distribution is right-skewed, with the highest frequencies in the mid-to-high range. This confirms that emotional exhaustion is a common experience among the participants. The presence of a long tail at the upper end indicates that a smaller subset is experiencing extreme burnout, which may significantly impact clinical performance and well-being. Different burnout phenotypes are revealed by the dual peaks:

- The first Peak (25): "Workload-driven exhaustion" - potentially addressable through workflow redesign;
- The second Peak (35): "Systemic burnout" - requiring organizational culture change.

The lower level at 30 points represents a critical threshold: once crossed, burnout becomes self-sustaining without intervention.

Together, these four subfigures illustrate a strong inverse relationship between resilience and emotional exhaustion. The Regression and between-group comparisons suggest that improving psychological resilience may be a viable strategy to mitigate burnout risk, while distributions show general trends and risk profiles within the sample.

Synthesis of Cross-Panel Insights:

The diagram below (Figure 8) depicts a proposed intervention framework based on the intersection of burnout severity and low resilience. It identifies two key risk groups: individuals with moderate and high burnout levels and resilience scores below 10. This indicates targeted actions, including workload adjustments and resilience training, as well as medical leave and multidisciplinary support.

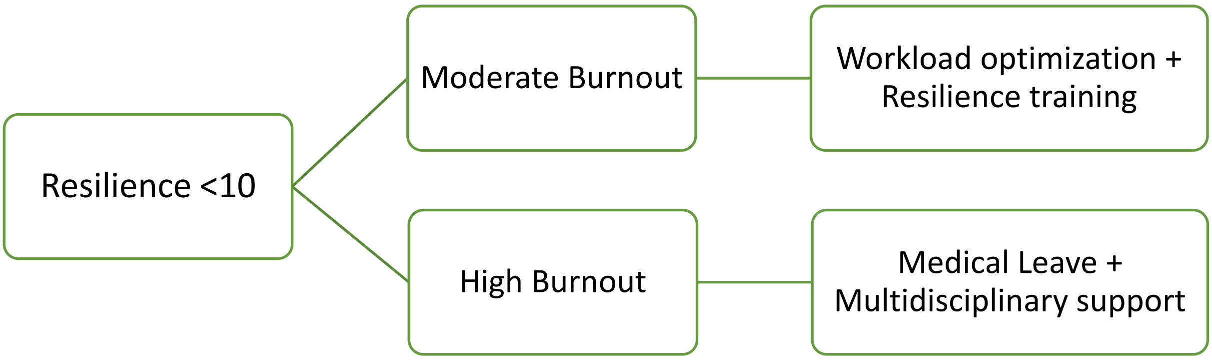

Figure 8 Intervention Targets Based on Burnout and Resilience Profiles

The analysis reveals two key implications at the organisational level. Firstly, the protective effect of resilience appears to diminish when exhaustion levels exceed 35 points, as suggested by the flattening of the trendline in Figure 7A. This indicates that resilience-building interventions may have limited impact among individuals already experiencing severe burnout, emphasising the need for earlier preventive action. Secondly, the unit-level distribution of exhaustion scores (Figure 7D) highlights specific clusters of high burnout. These clusters should be prioritised for targeted organisational responses, such as workload redistribution, psychological support or shift rotation policies.

As a continuation of the previous analysis, the Clinical Action Framework presented below (Table 4) outlines suggested solutions based on various combinations of burnout severity and resilience zones.

Table 4 Clinical Action Framework

Resilience Zone	**Burnout Level**	**Recommended Actions**
>14 (Top 12%)	Any level	Peer resilience mentors
10-14	Moderate (25-29)	Cognitive-affective skills training
7-10	High (>30)	Workload reduction + Therapy
<7	Severe (>35)	Medical leave + Intensive treatment

Table 4 illustrates how each panel offers complementary insights, with 7A revealing global patterns, 7B showing group differences and 7C/D exposing population distributions that inform targeted interventions.

6 CONCLUSIONS AND FUTURE RESEARCH

This thesis underscores the value of leveraging data analytics to detect and mitigate employee burnout in organisational settings. Through both a comprehensive literature review and a quantitative case study in a healthcare context, we found strong evidence that data-driven indicators can successfully flag early signs of burnout. In our analysis, workplace data such as hours worked and shift patterns, combined with psychometric measures (burnout inventory scores, resilience levels and situational anxiety), prove to be effective in identifying high-risk individuals. The incidence of moderate to high burnout among the healthcare professionals surveyed highlights that burnout is not a series of isolated personal cases but rather a widespread systemic problem. This reinforces the argument that organisations must take proactive, system-level measures. Our results showed that higher resilience scores are associated with significantly lower exhaustion, supporting the notion that reinforcing personal resilience can combat burnout. However, we also observed that beyond a certain threshold of burnout severity (exhaustion scores > 35), resilience alone offers diminishing protection. This implies that when burnout becomes extreme, more direct interventions (such as medical leave or workload redistribution) are necessary. In short, data analytics can provide an early warning system for burnout, allowing organisations to intervene before burnout becomes self-perpetuating. The research thus achieves its primary objective: demonstrating that by mining and analysing employee data, managers can gain actionable insights to maintain workforce well-being and productivity.

6.1 THEORETICAL AND PRACTICAL IMPLICATIONS

This work contributes to the emerging intersection of data science and occupational health literature. Theoretically, it validates frameworks which suggest that personal resources (e.g., resilience) and external stressors (e.g., pandemic fears and overwork) jointly influence burnout outcomes. Our findings align with prior models of burnout by quantifying these relationships in a real-world dataset. Notably, the strong negative correlation between resilience and exhaustion supports theories that resilience is a key mediator of stress. This confirms that employees with greater adaptive capacity handle job stressors more effectively. At the same time, the impact of contextual factors (like prolonged work hours and night shifts) on burnout levels is consistent with the traditional job-demands-resources paradigm, which emphasises that high job demands without adequate recovery lead to burnout. From a practical management perspective, the study offers a data-driven framework for intervention. Organisations should monitor workload metrics and employee feedback in real time, as this data can signal rising burnout risk. For example, our proposed Intervention Framework (Figure 8) illustrates targeted actions: employees experiencing high burnout and low resilience should receive priority support, such as reduced workloads, counselling or resilience training, while even those with moderate burnout can benefit from preventive measures like stress management workshops. Furthermore, identifying a small subset of “burnout-resistant” individuals (top ~12% resilience scores) suggests opportunities for peer mentoring programs, where these resilient employees can share coping strategies with colleagues. Overall, the integration of data analytics into HR practices could transform how organisations approach employee mental health: moving from reactive (addressing burnout after it manifests) to proactive (predicting and preventing burnout through continuous data monitoring).

6.2 LIMITATIONS

Despite its contributions, this research has several important limitations. First, the empirical study is based on a single dataset drawn from the healthcare sector in Iran during a pandemic. This specific context, front-line healthcare workers under COVID-19 stress, means the baseline burnout levels were extraordinarily high, which may not generalise to other industries or non-pandemic times. Cultural factors and healthcare system differences could also influence how burnout manifests and is reported, limiting the generalizability of our numeric thresholds or model. Second, due to data constraints, we were unable to test some of our initial hypotheses. In particular, physiological data from wearable devices (hypothesised to correlate with self-reported burnout) could not be obtained. This could change the predictive value of combining biometric data with organisational data. Similarly, hypothesis H4, which focuses on improvements in well-being following interventions that were inferred from literature rather than being measured, was not tested using a longitudinal design or an actual intervention. All analysis here is cross-sectional, capturing a snapshot in time. Causal interpretations should be made with caution. There is also the issue of self-report bias: the burnout and resilience measures rely on survey responses, which may be influenced by personal perception or reluctance to report extreme feelings. Lastly, the machine learning models we developed (Logistic Regression, Decision Tree and Random Forest) reached moderate accuracy but could likely be improved with more data or advanced features. We did not emphasise the model performance in the results, as the focus was on interpretability and insight. However, from a technical standpoint, there is room for refinement.

6.3 FUTURE RESEARCH

Building on this study, future research can take several fruitful directions. A clear next step is to incorporate multi-modal data, for example, collect real-time physiological indicators (heart rate variability, sleep patterns, etc.) via wearable sensors and combine them with organisational metrics and survey data. This would allow testing hypotheses H2 and H3 directly, potentially improving predictive accuracy for burnout and offering richer insights into the mind-body aspects of workplace stress. Longitudinal studies are also critical: by tracking employees over time, researchers could observe how early warning signs (like gradual increases in workload or stress biomarkers) translate into burnout outcomes down the line. Such studies would enable the development of time-sensitive predictive models (e.g., algorithms that alert management when an employee's risk profile has been trending upward). Another important path is to intervene and experiment based on analytics, implement targeted interventions (for instance, a resilience training program or a rotation policy to reduce consecutive night shifts) and then measure the subsequent changes in burnout levels and job performance. This would provide causal evidence for what preventive actions truly make a difference, addressing hypothesis H4 in practice. It would also be valuable to replicate and extend this research in diverse organisational contexts (different industries (tech, education, etc.), various cultural settings and both during crisis conditions and stable periods). This comparative approach would help validate whether the data indicators of burnout we identified (e.g., >60 hours/week or low resilience scores) hold universally or if different environments have different risk factors. Finally, future research should explore the ethical and privacy considerations of workplace analytics for mental health. While data-driven monitoring can be extremely useful, employees must trust that their data will be used to help them, not penalise them. Developing guidelines and tools (perhaps using anonymisation or aggregate trend reporting) could be important to implement these approaches responsibly. By pursuing these directions, scholars and practitioners can further evolve a data-informed strategy for combating burnout, one that safeguards employee well-being and fosters sustainable organisational performance.

BIBLIOGRAPHICAL REFERENCES

Ahorsu, D. K., Lin, C.-Y., Imani, V., Saffari, M., Griffiths, M. D., & Pakpour, A. H. (2020). The Fear of COVID-19 Scale: Development and Initial Validation. *International Journal of Mental Health and Addiction, 18*(4), 1537–1545. https://doi.org/10.1007/s11469-020-00270-8

American Psychological Association. (2023). Work in America: 2023 Workplace health and well-being report

Bakker, A. B., & Oerlemans, W. G. (2011). Subjective well-being in organisations.

Centers for Disease Control and Prevention. (2023). Vital Signs: Health Worker-Perceived Working Conditions and Symptoms of Poor Mental Health - Quality of Worklife Survey, United States, 2018-2022.

Dehdashti, A., Fatemi, F., Belji Kangarlou, M., & Paknazar, F. (2022). Dataset on working conditions and stressors and perceived mental health among Iranian healthcare workers (Version 2) [Data set]. Mendeley Data. https://doi.org/10.17632/ccdppxc6pb.2

Deloitte. (2022). Mental health and employers: The case for investment - pandemic and beyond.
https://www.deloitte.com/content/dam/assetszone2/uk/en/docs/industries/energy-resources-industrials/2023/deloitte-uk-mental-health-report-2022.pdf

Dyrbye, L. N., West, C. P., Satele, D., Boone, S., Tan, L., Sloan, J., & Shanafelt, T. D. (2014). Burnout among U.S. medical students, residents, and early career physicians. *Academic Medicine*, *89*(3), 443–451. https://doi.org/10.1097/ACM.0000000000000134

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139–183). North-Holland.

Kleijweg, J. H. M., Verbraak, M. J. P. M., & Van Dijk, M. K. (2013). The clinical utility of the Maslach Burnout Inventory in a clinical population. Psychological Assessment, 25(2), 435-441. https://doi.org/10.1037/a0031334

Maslach, C., & Leiter, M. P. (2016). Understanding the burnout experience: recent research and its implications for psychiatry. World Psychiatry, 15(2), 103-111. https://doi.org/10.1002/wps.20311

Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). Maslach Burnout Inventory Manual (3rd ed.). Consulting Psychologists Press

Matheny, M., Israni, S. T., Ahmed, M., & Whicher, D. (2019). Artificial intelligence in health care: The hope, the hype, the promise, the peril. Washington, DC: National Academy of Medicine, 10.

Morgantini, L. A., et al. (2020). Factors contributing to healthcare burnout during COVID-19. The Lancet, 396(10257), 1234-1238

Morgantini, L. A., Naha, U., Wang, H., Francavilla, S., Acar, Ö., Flores, J. M., ... & Weine, S. M. (2020). Factors contributing to healthcare professional burnout during the COVID-19 pandemic: A rapid turnaround global survey. *PloS one*, *15*(9), e0238217. https://doi.org/10.1371/journal.pone.0238217

Osborne, S. (2016). Employee engagement and organisational profitability. Walden Dissertations and Doctoral Studies. 3194. https://scholarworks.waldenu.edu/dissertations/3194

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. Doi: 10.1136/bmj.n71

Purvanova, R. K., & Muros, J. P. (2010). Gender differences in burnout: A meta-analysis. *Journal of Vocational Behaviour*, *77*(2), 168–185. https://doi.org/10.1016/j.jvb.2010.04.006

Salvagioni, D. A. J., Melanda, F. N., Mesas, A. E., González, A. D., Gabani, F. L., & Andrade, S. M. D. (2017). Physical, psychological and occupational consequences of job burnout: A systematic review of prospective studies. PloS one, 12(10), e0185781 https://doi.org/10.1371/journal.pone.0185781

Schaufeli, W. B., & Taris, T. W. (2014). A critical review of the Job Demands-Resources Model: Implications for improving work and health. In G. F. Bauer & O. Hämmig (Eds.), *Bridging Occupational, Organisational and Public Health* (pp. 43–68). Springer. https://doi.org/10.1007/978-94-007-5640-3_4

Schaufeli, W. B., & Taris, T. W. (2005). The conceptualization and measurement of burnout: Common ground and worlds apart. Work & Stress, 19(3), 256–262. https://doi.org/10.1080/02678370500385913

World Health Organisation. (2019). Burn-out is an "occupational phenomenon": International Classification of Diseases (ICD-11). https://www.who.int/news/item/28-05-2019-burn-out-an-occupational-phenomenon-international-classification-of-diseases

World Health Organisation. (2023). Stress. World Health Organisation. https://www.who.int/news-room/questions-and-answers/item/stress

World Health Organisation. (2024). ICD-11 for Mortality and Morbidity Statistics. Retrieved from https://icd.who.int/browse/2024-01/mms/en#129180281

APPENDIX A - Questionnaire

Semnan University of Medical Science
and Health Services

Date:

Questionnaire ID:

Questionnaires on personal and work characteristics, mental health, and task load and fear of Covid-19

Part 1: Demographic features of the study participants:

Demographic Characteristics				
Gender	☐ Female		☐ Male	
Marital Statues	☐ Single		☐ Married	
Employment Statues	☐ Permanent	☐ Temporary	☐ Informal	
Working Hours	☐ 7h	☐ 8h	☐ 10h	☐ 12h
Training on safe work procedures	☐ Yes		☐ No	
Evaluation of safety and health at work environment	☐ Acceptable		☐ Unacceptable	

Part 2: Perceived mental health domain: Occupational exhaustion, depersonalization, personal accomplishment.

Indicate how frequently the following statements apply to you and add the points indicated on top of the respective box:

0 = Never

1 = At least a few times a year

2 = At least once a month

3 = Several times a month

4 = Once a week

5 = Several times a week

6 = Every day

Questions	Never 0	At least a few times a year 1	At least once a month 2	Several times a month 3	Once a week 4	Several times a week 5	Every day 6
1. I feel emotionally exhausted because of my work							
2. I feel worn out at the end of a working day							
3. I feel tired as soon as I get up in the morning and see a new working day stretched out in front of me							
4. Working with people the whole day is stressful for me							
5. I feel burned out because of my work							
6. I feel frustrated by my work							
7. I get the feeling that I work too hard							
8. Being in direct contact with people at work is too stressful							
9. I feel as if I'm at my wits' end							
10. I get the feeling that I treat some clients/colleagues impersonally, as if they were object							
11. I have become more callous to people since I have started doing this job							
12. I'm afraid that my work makes me emotionally harder							
13. I'm not really interested in what is going on with many of my colleagues							
14. I have the feeling that my colleagues blame me for some of their problems							
15. I can easily understand the actions of my colleagues/supervisors							
16. I deal with other people's problems successfully							
17. I feel that I influence other people positively through my work							
18. I feel full of energy							
19. I find it easy to build a relaxed atmosphere in my working environment							

20. I feel stimulated when I been working closely with my colleagues							
21. I have achieved many rewarding objectives in my work							
22. In my work I am very relaxed when dealing with emotional problems							

Overall score for occupational exhaustion (EE)
Add together the answers to questions 1-9

Occupational exhaustion	EE < 17	EE 18 - 29	EE > 30
	Low degree	Moderate degree	High degree

Overall score for depersonalization / loss of empathy (DP)
Add together the answers to questions 10-14

Depersonalization	DP < 5	DP 6 - 11	DP > 12
	Low degree	Moderate degree	High degree

Overall score personal accomplishment assessment (PA)
Add together the answers to questions 15-22

Personal Accomplishment	PA < 33	PA 34 - 39	PA > 40
	Low degree	Moderate degree	High degree

<u>Degree of burnout</u>

EE	Occupational exhaustion (burnout) is typically connected to a relationship with work that is perceived as difficult, tiring, stressful... Maslach sees this as different from depression, as it is likely that the symptoms of burnout would be reduced during holidays.
DP	Depersonalization or loss of empathy is characterized by a loss of regard for others (clients, colleagues...), and by keeping a greater emotional distance, which is expressed through cynical, derogatory remarks, and even callousness.
PA	The personal accomplishment assessment is a feeling that acts as a "safety valve" and contributes to bringing about a balance if occupational exhaustion and depersonalization

	occur. It ensures fulfilment in the workplace and a positive view of professional achievements.

Part 3: NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Very Low — Very High

Effort — How hard did you have to work to accomplish your level of performance?

Very Low Very High

How insecure, discouraged, irritated, stressed, and annoyed were you?

Frustration

Very Low Very High

Part 4: Fear of COVID-19 Scale

1. Do you fear of catching coronavirus infection?
2. Do you worry about taking it home to family?
3. Do you fear for the future?
4. Do you fear of death from Covid-19?

Scoring

The participants indicate their level of agreement with the statements using a four- item Likert-type scale. Answers included “no” to “high”. The scores were from 4 to 16. The total score was classified as negligible (≤8), moderate (9–12), and high fear of the pandemic.

APPENDIX B: CODING FOR THE FIGURES

The entire analytical workflow was implemented in Python 3.9 using PyCharm Community Edition 2023.2.1 as the integrated development environment (IDE). Core scientific libraries included pandas for data manipulation, scikit-learn for machine learning algorithms and seaborn for statistical visualization. The complete codebase is organized into modular scripts with the following structure:

- Data Pipeline (preprocessing.py): Automated handling of missing values, categorical encoding (e.g., shift types) and feature scaling (MinMaxScaler);
- Exploratory Analysis (eda.py): Generation of Figures 2-7 through statistical plotting functions;
- Model Training (models.py): Implementation of Logistic Regression, Decision Tree and Random Forest;
- Validation Suite (validation.py): Calculation of performance metrics (Accuracy, Precision, F1-Score) using 5-fold cross-validation

```
# ========================

# 1. IMPORT LIBRARIES

# ========================

import os

import matplotlib.pyplot as plt

import pandas as pd

import seaborn as sns

from scipy.stats import ttest_ind

from sklearn.linear_model import LogisticRegression

from sklearn.model_selection import train_test_split

from sklearn.preprocessing import MinMaxScaler


# Set plotting style

sns.set_style("whitegrid")

plt.rcParams['figure.dpi'] = 100

plt.rcParams['savefig.dpi'] = 300
```

```
# ========================
# 2. CHECK DIRECTORY AND LOAD DATA
# ========================
print("Current Working Directory:", os.getcwd())
print("\nFiles in this directory:")
print(os.listdir())

try:
    df = pd.read_csv("mental health Dataset.csv")
    print("\nDataset loaded successfully! First 5 rows:")
    print(df.head())

    # Check unique values in key columns
    print("\nUnique values in 'Shift':", df['Shift'].unique())
    print("Unique values in 'Sex':", df['Sex'].unique())
    print("Unique values in 'Married':", df['Married'].unique())

    # Map shift values to descriptive labels
    shift_mapping = {
        1: 'Day',
        2: 'Night',
        3: 'Rotating',
        4: 'Evening',
        5: 'Mixed',
        6: 'On-call',
```

```
        7: 'Other'}
    df['Shift'] = df['Shift'].map(shift_mapping).fillna('Unknown')

    # Map gender values
    gender_mapping = {
        1: 'Male',
        2: 'Female'}
    df['Sex'] = df['Sex'].map(gender_mapping).fillna('Other')

    # Map marital status
    married_mapping = {
        1: 'Married',
        2: 'Single',
        3: 'Divorced',
        4: 'Widowed'}
    df['Married'] = df['Married'].map(married_mapping).fillna('Other')

except Exception as e:
    print("\nError loading the dataset:", e)
    exit()

# =======================
# 3. DATA PREPROCESSING
# =======================

# Convert burnout columns to numeric
```

```
burnout_columns = [f'q{i}Burnout' for i in range(1, 20)]
df[burnout_columns] = df[burnout_columns].apply(pd.to_numeric, errors='coerce')

# Calculate burnout scores
df['Exhaustion'] = df[[f'q{i}Burnout' for i in range(1, 10)]].sum(axis=1)
df['Depersonalization'] = df[[f'q{i}Burnout' for i in range(10, 15)]].sum(axis=1)
df['Personal_Achievement'] = df[[f'q{i}Burnout' for i in range(15, 20)]].sum(axis=1)

# Classify exhaustion levels
df['Exhaustion_Level'] = pd.cut(
    df['Exhaustion'],
    bins=[0, 17, 29, 54],
    labels=['Low', 'Moderate', 'High'])

# Convert COVID fear columns to numeric and calculate score
covid_fear_columns = ['q1Anexity', 'q2Anexity', 'q3Anexity', 'q4Anexity']
df[covid_fear_columns] = df[covid_fear_columns].apply(pd.to_numeric, errors='coerce')
df['COVID_Fear_Score'] = df[covid_fear_columns].sum(axis=1)

# ========================
# 4. MINMAX SCALING
# ========================
scaler = MinMaxScaler()
df_scaled = pd.DataFrame(
    scaler.fit_transform(df[['Exhaustion', 'COVID_Fear_Score']]),
    columns=['Exhaustion', 'COVID_Fear_Score'])
```

```
# ========================
# 5. STATISTICAL ANALYSIS
# ========================
# T-test between shifts (Day vs Night)
df_clean = df.dropna(subset=['Shift', 'Exhaustion'])
group1 = df_clean[df_clean['Shift'] == 'Day']['Exhaustion']
group2 = df_clean[df_clean['Shift'] == 'Night']['Exhaustion']

if not group1.empty and not group2.empty:
    print("\nT-test between shifts (Day vs Night):")
    t_result = ttest_ind(group1, group2)
    print(f"T-statistic: {t_result.statistic:.4f}, P-value: {t_result.pvalue:.4f}")
else:
    print("\nInsufficient data for t-test between shifts")

# ========================
# 6. DESCRIPTIVE ANALYSIS AND VISUALIZATION
# ========================
df.dropna(inplace=True)

# Frequency of exhaustion levels
print("\n=== Burnout Levels Distribution ===")
print(df['Exhaustion_Level'].value_counts(normalize=True).mul(100).round(1))

# Plot 1: Burnout Levels Distribution
```

```
plt.figure(figsize=(10, 6))
ax = sns.countplot(data=df, x='Exhaustion_Level', order=['Low', 'Moderate', 'High'],
                palette='viridis', hue='Exhaustion_Level', legend=False)
plt.title("Occupational Burnout Levels Distribution", fontsize=16)
plt.xlabel("Burnout Level", fontsize=12)
plt.ylabel("Count", fontsize=12)

total = len(df)
for p in ax.patches:
    percentage = f'{100 * p.get_height() / total:.1f}%'
    ax.annotate(percentage, (p.get_x() + p.get_width() / 2, p.get_height()),
            ha='center', va='bottom', fontsize=10)
plt.tight_layout()
plt.savefig('burnout_levels.png')
plt.show()

# Plot 2: Age Distribution
plt.figure(figsize=(10, 6))
sns.histplot(df['Age'], bins=20, kde=True, color='royalblue')
plt.title("Age Distribution of Healthcare Workers", fontsize=16)
plt.xlabel("Age", fontsize=12)
plt.ylabel("Frequency", fontsize=12)
plt.tight_layout()
plt.savefig('age_distribution.png')
plt.show()
```

```
# Plot 3: Burnout by Gender
plt.figure(figsize=(10, 6))
sns.boxplot(data=df, x='Sex', y='Exhaustion', hue='Sex',
        palette='coolwarm', legend=False)
plt.title("Burnout Levels by Gender", fontsize=16)
plt.xlabel("Gender", fontsize=12)
plt.ylabel("Exhaustion Score", fontsize=12)
plt.tight_layout()
plt.savefig('burnout_gender.png')
plt.show()

# Plot 4: Burnout vs Working Hours (WPW = Weekly Working Hours)
plt.figure(figsize=(10, 6))
sns.scatterplot(data=df, x='WPW', y='Exhaustion', hue='Shift',
           palette='dark:salmon_r', s=80, alpha=0.8)
plt.title("Burnout Levels vs Weekly Working Hours", fontsize=16)
plt.xlabel("Weekly Working Hours (WPW)", fontsize=12)
plt.ylabel("Exhaustion Score", fontsize=12)
plt.legend(title='Work Shift', loc='upper right')
plt.tight_layout()
plt.savefig('burnout_vs_hours.png')
plt.show()

# Plot 5: Correlation Matrix
plt.figure(figsize=(10, 6))
correlation_matrix = df[['Exhaustion', 'COVID_Fear_Score', 'q1Resilence']].corr()
```

```
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', vmin=-1, vmax=1)

plt.title("Correlation Matrix of Key Variables", fontsize=16)

plt.tight_layout()

plt.savefig('correlation_matrix.png')

plt.show()


# ========================

# 7. RESILIENCE ANALYSIS

# ========================

resilience_columns = [f'q{i}Resilence' for i in range(1, 6)]

df[resilience_columns] = df[resilience_columns].apply(pd.to_numeric, errors='coerce')

df['Resilience_Score'] = df[resilience_columns].sum(axis=1)


# Plot 6: Resilience vs Burnout

plt.figure(figsize=(12, 8))


# Create a 2x2 grid of plots

plt.subplot(2, 2, 1)  # Top left - Scatter plot with regression line

sns.regplot(x='Resilience_Score', y='Exhaustion', data=df,

        scatter_kws={'alpha': 0.3, 'color': 'steelblue'},

        line_kws={'color': 'red', 'linewidth': 2.5})

plt.title("Relationship: Resilience vs Burnout", fontsize=14)

plt.xlabel("Resilience Score", fontsize=12)

plt.ylabel("Exhaustion Score", fontsize=12)


plt.subplot(2, 2, 2)  # Top right - Boxplot by burnout level
```

```
sns.boxplot(x='Exhaustion_Level', y='Resilience_Score', data=df,
            palette='Set2', order=['Low', 'Moderate', 'High'])
plt.title("Resilience by Burnout Level", fontsize=14)
plt.xlabel("Burnout Level", fontsize=12)
plt.ylabel("Resilience Score", fontsize=12)

plt.subplot(2, 2, 3)  # Bottom left - Distribution of resilience scores
sns.histplot(df['Resilience_Score'], bins=15, kde=True, color='seagreen')
plt.title("Distribution of Resilience Scores", fontsize=14)
plt.xlabel("Resilience Score", fontsize=12)
plt.ylabel("Frequency", fontsize=12)

plt.subplot(2, 2, 4)  # Bottom right - Distribution of exhaustion scores
sns.histplot(df['Exhaustion'], bins=15, kde=True, color='purple')
plt.title("Distribution of Exhaustion Scores", fontsize=14)
plt.xlabel("Exhaustion Score", fontsize=12)
plt.ylabel("Frequency", fontsize=12)

plt.suptitle("Comprehensive Analysis: Resilience vs Burnout", fontsize=18, y=0.98)
plt.tight_layout(pad=3.0)
plt.savefig('resilience_vs_burnout.png')
plt.show()

# =======================
# 8. LOGISTIC REGRESSION MODEL
# =======================
```

```
# Convert burnout levels to numeric codes
df['Exhaustion_Level'] = df['Exhaustion_Level'].astype('category')
y = df['Exhaustion_Level'].cat.codes

X = df[['Age', 'WPW', 'Resilience_Score', 'COVID_Fear_Score']]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

model = LogisticRegression(max_iter=1000)
model.fit(X_train, y_train)
print(f"\nModel Accuracy: {model.score(X_test, y_test):.2%}")

# =======================
# 9. RESULTS TABLES
# =======================
# Table 1: Descriptive Statistics
desc_stats = df[['Exhaustion', 'COVID_Fear_Score', 'Resilience_Score']].describe()
print("\nTable 1 - Descriptive Statistics:\n", desc_stats)

# Table 2: Average Burnout by Shift
burnout_shift = df.groupby('Shift')['Exhaustion'].mean().round(1)
print("\nTable 2 - Average Burnout by Work Shift:\n", burnout_shift)

# Table 3: Key Correlations
key_correlations = df[['Exhaustion', 'Depersonalization', 'COVID_Fear_Score', 'Resilience_Score']].corr().round(2)
print("\nTable 3 - Key Variable Correlations:\n", key_correlations)
```

APPENDIX C: Generative AI ASSISTANCE

Tasks done with GenAI help	GenAI Tools	Specific Application
Grammar check	Grammarly; Quill Bot	Grammar checking and text polishing
Data visualisation coding	ChatGPT; DeepSeek	Generation and debugging of figure code
Conceptual development	ChatGPT; DeepSeek	Brainstorming and problem-solving support
Literature research	Consensus	Academic paper discovery and synthesis

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa