MENTAL STRESS AND OVERLOAD DETECTION

MENTAL STRESS AND OVERLOAD DETECTION FOR OCCUPATIONAL SAFETY

By SAHEL ESKANDAR, M.A.Sc., B.A.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Doctor of Philosophy in Engineering

Doctor of Philosophy in Engineering (2022)

McMaster University, Hamilton, Ontario

TITLE: Mental Stress and Overload Detection for Occupational Safety

AUTHOR: Sahel Eskandar, M.A.Sc., B.A.Sc (McMaster University)

SUPERVISOR: Dr. Saiedeh Razavi

NUMBER OF PAGES: xv, 130,

ABSTRACT

Stress and overload are strongly associated with unsafe behaviour, which motivated various studies to detect them automatically in workplaces. This study aims to advance safety research by developing a data-driven stress and overload detection method. An unsupervised deep learning-based anomaly detection method is developed to detect stress. The proposed method performs with convolutional neural network encoder-decoder and long short-term memory equipped with an attention layer. Data from a field experiment with 18 participants was used to train and test the developed method. The field experiment was designed to include a pre-defined sequence of activities triggering mental and physical stress, while a wristband biosensor was used to collect physiological signals. The collected contextual and physiological data were pre-processed and then resampled into correlation matrices of 14 features. Correlation matrices are used as an input to the unsupervised Deep Learning (DL) based anomaly detection method. The developed method is validated, offering accuracy and F-measures close to 0.98. The technique employed captures the input data attributes correlation, promoting higher interpretability of the DL method for easier comprehension. Over-reliance on uncertain absolute truth, the need for a high number of training samples, and the requirement of a threshold for detecting anomalies are identified as shortcomings of the proposed method. To overcome these shortcomings, an Adaptive Neuro-Fuzzy Inference System (ANFIS) was designed and developed. While the ANFIS method did not improve the overall accuracy, it outperformed the DL-based method in detecting anomalies precisely. The overall performance of the ANFIS method is better than the DL-based method for the anomalous class, and the method results in lower false alarms. However, the DL-based method is suitable for circumstances where false alarms are tolerated.

Keywords: Worker Safety; Stress and Overload; Data-Driven Health Monitoring; Wearable Sensors; Unsupervised Learning; Adaptive Neuro-Fuzzy Inference System

ACKNOWLEDGEMENT

I would like to express my most profound appreciation to my supervisor, Dr. Saiedeh Razavi, who supported me throughout this journey. This research would only happen with her insight, knowledge, and mentorship. I am grateful to my committee members, Dr. Zoe Li and Dr. Qiyin Fang, for their support and helpful advice.

I also want to thank my family for their love and encouragement. From the 10,000 km distance, my father and mother's support in the hardest moments made this happen. I appreciate having a passionate and ambitious sister and brother who made me keep the standards high and push through every step.

Many thanks to the Natural Sciences and Engineering Research Council of Canada (NSERC) for the support of this research under NSERC Discovery Accelerator Supplement Grant No: RGPAS-2018-522618 and NSERC Discovery Grant No: RGPIN-2018-05434.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACC	Accelerometer
ANFIS	Adaptive Neural Fuzzy Inference System
ARMA	Auto Regressive Moving Average
ARIMA	Auto Regressive Integrated Moving Average
BP	Blood Pressure
BVP	Blood Volume Pulse
CNN	Convolutional Neural Network
ConvLSTM	Convolutional Long-Short Term Memory
CPS	Cyber Physical Systems
DL	Deep Learning
EDA	Electrodermal Activity
EEG	Electrocochleography
FIS	Fuzzy Inference System
FN	False Negative
FP	False Positive
GSR	Galvanic Skin Response
HiLCPS	Human in the Loop Cyber Physical Systems
HR	Heart Rate
IBI	Inter Beat Interval
LOWESS	Locally Weighted Scatterplot Smoothing
LSTM	Long Short-Term Memory

MF	Membership Function
ML	Machine Learning
MSCRED	Multi-Scale Convolutional Recurrent Encoder-Decoder
MREB	McMaster Research Ethics Board
PPE	Personal Protective Equipment
PPG	Photoplethysmography
RR	Respiration Rate
ST	Skin Temperature
SVM	Support Vector Machine
TEMP	Temperature
TN	True Negative
TP	True Positive

DECLARATION OF ACADEMIC ACHIEVEMENT

"I, Sahel Eskandar, declare that the Ph.D. thesis entitled mental stress and overload detection for occupational safety is not more than 300 pages long, including quotes and exclusive of tables, figures, bibliography, references, and footnotes. This thesis contains no material submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my work".

Signature …….. Date …….

Chapter 1 Introduction

1.1 Chapter Overview

This chapter presents the research problem statement as the primary stage of the study. It starts by addressing Motivation and Background in section 1.2 and defining research objectives in section 1.3. Research methodology and scopes are described in section 1.4, followed by thesis organization in section 1.5.

1.2 Motivation and Background

Due to the significant rate of accidents and fatalities in construction sites, safety remains a considerable concern in construction environments. Despite the advancement in research, developments, and regulations for safety in construction, the construction industry remains the most fatal among other sectors in Canada (AWCBC/ACATC, 2020). Canada Work Injury, Disease, and Fatality Statistics (Fig. 1) showed that fatal accidents in construction accounted for almost 20 percent of the reported workers' fatalities in 2020 (AWCBC/ACATC, 2020). Despite the start of the Covid 19 pandemic in 2020 and the extraordinary risk of working in the healthcare industry, 29 fatalities were reported in the healthcare and social assistance industry (Fig. 1). In the same year, 193 cases were reported for the construction industry, which is more than six times the healthcare reported fatalities cases (AWCBC/ACATC, 2020).

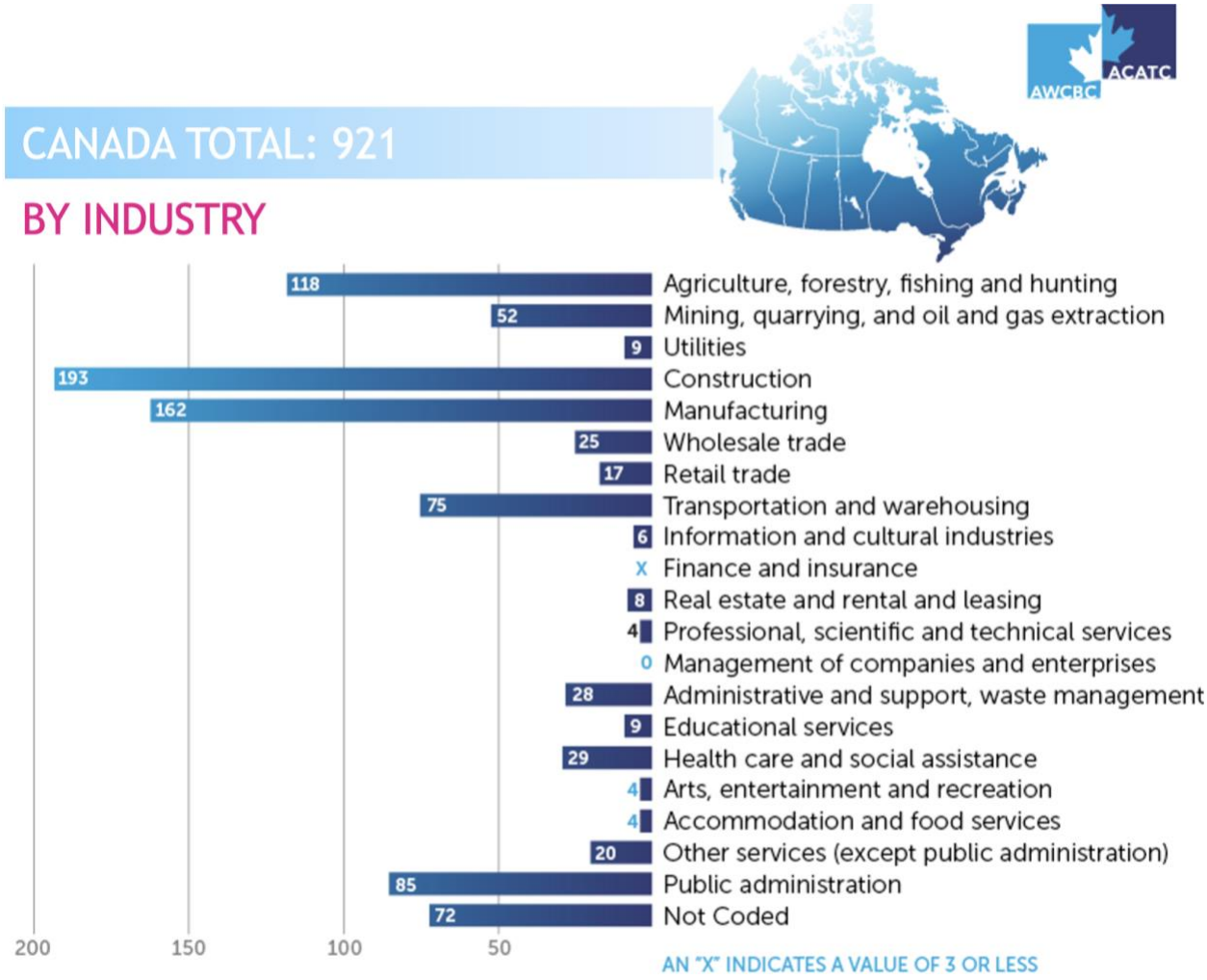

Fig. 1. 2020 Fatalities from the Association of Workers' Compensation Boards of Canada (AWCBC/ACATC, 2020)

Accidents in construction are caused by reasons such as lack of training, unsafe equipment, unsafe site conditions, unsafe methods, poor management, as well as human and social factors (Abdelhamid & Everett, 2000a; Adams et al., 2014; Khosravi et al., 2015; Toole, 2002). Multiple models have been developed for tracing the root causes of an accident, and the Accident Root Cause Tracing Model (Abdelhamid & Everett, 2000b) identified unsafe acts, unsafe conditions, or both as direct causes of accidents. Among conditions and causal factors leading to accidents,

dangerous human behaviour and human error are identified as the leading cause of accidents (Chua & Goh, 2004; HEINRICH, 1950; Reason, 1991).

As unsafe human behaviour is widely identified as the most frequent cause of accidents, most studies in construction safety focused on studying human factors leading to dangerous behaviour (Eskandar et al., 2019). Many construction safety studies focused on construction workers' cognitive processes as unsafe behaviour is driven by failures at some stages of human cognition (Hasanzadeh et al., 2017; Reason, 1991). From a human cognition perspective, understanding how individuals perceive information and decide to take action can uncover why an unsafe act may follow (Fang et al., 2016).

A significant factor influencing cognition is the lack of situation awareness (SA), as Endsley described a detailed relation of SA to the human cognition process (Endsley 1988c, 1988a). Also, (Bedny & Meister, 1999) considered SA part of the cognitive process of dynamic systems. Precisely, in the construction industry, due to the excessive workload in a dynamic and complex work setting, the lack of SA has an evident impact on human behaviour (Hasanzadeh et al., 2016, 2017). Higher cognitive loads increase muscle activity and impact task completion time (Biondi et al., 2021). Workers' lack of attention could prevent them from identifying an incident that leads to an accident (Manchi et al., 2013).

Moreover, due to the physically demanding nature of construction activities, many workers go beyond the accepted and safe physiological level for manual work (Abdelhamid & Everett, 2002). As Lee et al. (2017) stated, one of the contributing features to unsafe behaviours in a construction zone is workers' physical status, which reflects the safety and productivity of the work. Therefore,

it is essential to consider workers' physiological situation in research and development for construction safety.

From a physiological and neurological perspective, stress and overload are strongly associated with unsafe behaviour, which motivated the current study to detect stress and overload in the workplace. Stress, as a biological response to the pressures on the nervous system (Deary, 1996), can be measured through subjective tests and questionnaires of individual responses (Reiss et al., 1986; Wittchen & Boyer, 1998).

Stress can be detected through variations in physiological features such as heartbeats, body temperature, and respiration rate (Harari et al., 2016; Kothgassner et al., 2016; Vrijkotte et al., 2000). Due to the enhancements in biosensors and real-time data collection, many recent studies focused on the physiological impacts of stress on the body to develop a stress detection model (Can, Arnrich, et al., 2019; Can, Chalabianloo, et al., 2019; Nath et al., 2020; Yadav et al., 2020). In identifying research gaps by reviewing the previous literature, it was revealed that the existing stress detection studies are mainly focused on and differ from each other in (1) data collection method (e.g., wearable sensors, pupil tracking, questionnaire); (2) selected features and variety of physiological signals; (3) the stress-inducing approach and design of the field experiment (e.g., data collection location and designed activities); and (4) learning algorithms.

The review of the related literature revealed that the current stress detection models based on physiological signals could benefit from further situational or contextual input data, such as the subjects' basic information. Another identified gap is the inadequate stress-inducing approaches considered in field experiments in stress detection research. An important finding was that no applied anomaly detection algorithms had been used so far to detect stress, while unsafe stress

levels can be considered anomalies. This research differs from the existing literature in input features, various stress-inducing activities in the field experiment, and the use of multivariate anomaly detection algorithms. It aims to train a stress detection algorithm that provides an interpretable method for easier comprehension.

1.3 Research Objectives

This research aims to advance safety research by developing intelligent stress and overload detection model. While the goal of this study is to detect abnormal states of occupational stress and overload, the objectives of this study are:

1) Design and implementation of a field experiment aiming to collect a range of data, including physiological and contextual data, while inducing stress;
2) Design, develop, and evaluate anomaly detection method(s) for stress and overload detection;
3) Align the model with the long-term objectives and future application in the construction industry as a potential human in the loop cyber-physical system.

1.4 Research Methodology and Scope

To address the research objective of detecting anomalous states of stress and overload, designing an in-person field experiment to provide data for learning, testing and validating the outcomes was needed. Therefore, a field experiment was designed and implemented, in which contextual data and physiological were collected. Using the collected data, the anomaly detection methods are designed, developed, and evaluated for workplace stress and overload using (1) Deep Learning-

Based, and (2) Adaptive Neuro-Fuzzy Inference System anomaly detection method, discussed in Chapter 5 and Chapter 6.

This research was conducted within the following scope to achieve the defined objectives.

- The field experiment for this study was conducted using sensors that can be worn outside clinical environments (i.e., non-invasive wearable devices)
- The stress detection model functions based on contextual and physiological data collected using a wristband with a frequency equal to or below one record per second.
- The stress detection model doesn't record participants' insight and confirmation (perceived stress level through questionnaire).
- Abnormal states of stress, but not the types of activities, are detected.
- The stress detection model explores the adaptivity of application for construction workers through wearing personal protective equipment and working in a noisy and dynamic environment.

1.5 Thesis Organization

The chapters of this thesis are organized as follows. Chapter 1 discusses the problem statement as the primary stage of the research. It started with discussions on motivation and background in section 1.2, defining research objectives in section 1.3, describing research methodology and scope in section 1.4, and thesis organization in section 1.5.

Following the problem statement, a comprehensive literature review was conducted and presented in Chapter 2., aligned with the identified objectives presented in Chapter 1. First, stress and overload impact on unsafe behaviour is investigated in section 2.2. Then, Physiological Data

Associated with Stress and Overload is studied in section 2.3 to recognize the physiological data that can reveal anomalous stress states in humans. In section 2.4, we went through different Stress Detection Methods and Algorithms that were used in previous research studies. After selecting the desired system for stress detection, various aspects of the algorithm that uses Multivariant Time Series Anomaly Detection are explored in section 2.5. Finally, aligned with the long-term objectives and future application of the model in the construction industry, Cyber-Physical Systems for Construction Safety is studied in section 2.6.

In Chapter 3, the project scope, objectives, and methodology for two subjects, (1) anomaly detection and (2) fuzzy inference system, are presented. Chapter 4 presents the data collection and pre-processing stages. Details on the field experiment, such as designed activities to perform by participants, equipment used for recording data, location, and timeline of the in-person experiment, are presented in 4.2. Section 4.3 focuses on the three different types of collected data, namely (1) basic-level data, (2) situation-level data, and (3) signal-level data. Data pre-processing, including signal-level data normalization, resampling, and standardization, are reported in section 4.4. Conducting an in-person field experiment requires special measures that are explained in section 4.5 which includes ethics, health and safety standards of the field experiment.

The design and implementation of the anomaly detection method are presented in Chapter 5. This method is based on the correlation matrices given in 5.2. The method’s structure, training process, and validation results for this method are shown in sections 5.3, 5.4, and 5.5. Section 5.6 concludes the method's results and addresses the limitations that motivate the study presented in chapter 6.

Adaptive Neuro-Fuzzy Inference System (ANFIS) as the second method in response to the limitation of the first method is shown in Chapter 6. This chapter starts by presenting some key

background information on the adopted method in section 6.2. Section 6.3 follows the proposed method, with details in section 6.4. Experimental results and analysis are reported in section 6.5, followed by concluding remarks in section 6.6.

The final chapter presents the research summary in section 7.2, followed by research contributions in section 7.3. Limitations and future work related to (1) experiment design and data collection, (2) stress detection using unsupervised deep learning, and (3) stress detection using ANFIS are presented in sections 7.4.1, 7.4.2, and 7.4.3. Method's comparison in section 7.5 presents the comparison and discussion regarding methods' efficiency and training cost and methods' performance metrics in sections 7.5.1 and 7.5.2. The final chapter is closed by concluding remarks in section 7.6. References used in this research are delivered after the last chapter. This thesis contains three Jupyter Notebooks (Eskandar, 2022a, 2022b, 2022c) uploaded in the GitHub repository name Sahel-Eskandar/PhD_Thesis, which include Pre-Processing (Eskandar, 2022c), Deep-Learning (Eskandar, 2022b), and ANFIS (Eskandar, 2022a) python codes used for this research with comments and outputs.

Chapter 2 Research Background

2.1 Chapter Overview

This chapter continues the preliminary stage presented in Chapter 1 through the following outline. First, as discussed in the previous chapter, mental stress and overload directly impact unsafe behaviour. Therefore, Stress and Overload Impact on Unsafe Behaviour is investigated in section 2.2. Then, preceding similar publications have been studied to identify essential features leading to and correlating with stress. Section 2.3, named Physiological Data Associated with Stress and Overload, helps this research to recognize the biological data that can reveal stressful moments in the human body. Afterward, section 2.4 presents different ways of stress detection and their algorithms, and the pros and cons of their application for construction settings are discussed. This section helps us to narrow down the method, which allows us to reach the research objectives. After studying stress detection algorithms, the anomaly detection method is selected for detecting moments of stress. Different algorithms that use Multivariant Time Series Anomaly Detection are explored in section 2.5. Aligned with the long-term objectives and future application of the model in the construction industry, Cyber-Physical Systems for Construction Safety are studied in section 2.6. Finally, research gaps are presented in section 2.7.

2.2 Stress and Overload Impact on Unsafe Behaviour

Physical and psychological stress are among the contributing features leading to unsafe behaviours (Leung et al., 2010, 2016). Examples of stress factors in workplaces such as construction include (1) physical stressors (like noise, vibration, lighting, boredom, fatigue, cold or heat); and (2) social and psychological stressors (like fear, uncertainty, anxiety, mental overload, and time pressure) (Choudhry et al., 2007; Goldenhar et al., 2003; Langdon & Sawang, 2018; Leung et al., 2016; Shakerian et al., 2021). The dynamic and ever-changing nature of the work and environment in the construction industry also contributes to workers' mental overload as an indispensable stressor (Love et al., 2010).

2.3 Physiological Data Associated with Stress and Overload

Physiological data collection has become a ubiquitous task, using a vast network of wearable sensors, enabling unobtrusive real-time data collection. Currently, the human body's vital signs can be quickly and practically recorded through wearable biosensors and health gadgets (e.g., smart watches, earbuds, and headsets). Many researchers used such sensors to measure specific physiological conditions to study factors that affect individual neurological statuses, such as stress (Jebelli et al., 2019), sleep deprivation (Powell & Copping, 2010), fatigue (Abdelhamid & Everett, 2000c; Abuwarda et al., 2023), and social aspects (Choudhry et al., 2009; Hasan et al., 2022; Yadav et al., 2020; Zokaei et al., 2020).

As Gedam and Paul (2021), after retrieving data from 9334 papers, presented, in detecting stress using wearable sensors, one or more psychophysiological parameters such as Heart Rate (HR),

Skin Temperature (ST), Galvanic Skin Response (GSR), Respiration Rate (RR), Accelerometer (ACC), Blood Pressure (BP), Electrocochleography (EEG) have been used.

Among different biological measuring methods that could reflect stress and mental overload, Electroencephalogram (EEG) sensors have been commonly applied in many studies (Chen et al., 2017; Jebelli et al., 2018; Wang et al., 2017). EEG is a valuable source in identifying brain activities by measuring the brain neurons' electrical activities using electrodes positioned on the scalp. However, there are limitations in its applications for extensive physical activities, such as those in Construction. EEG signals are sensitive to face and body movements (e.g., eye blinks), making the method impractical for construction safety applications (Lew et al., 2012; Teplan, 2002; D. Wang et al., 2017). Similarly, cortisol level measurement, as an indicator of stress hormone secretion, is considered impractical for free-moving conditions. Considering the above limitations, viable biosensors that could detect and reflect the stress in workplaces such as construction are Photoplethysmography (PPG), Electrodermal Activity (EDA), and peripheral skin temperature (ST) (Jebelli et al., 2019).

Measurement of cardiovascular parameters using a photoplethysmography optical sensor (i.e., PPG) has been vastly used in the past. PPG can record real-time data in body relaxation and stress stimulation situations and precisely recall stress (Heo et al., 2021; Přibil et al., 2020). On the other hand, Electrodermal activity (EDA) is an indicator of sympathetic nervous system initiation, known as one of the specific and practical symptoms of stress arousal (Pakarinen et al., 2019). Spike in stress level activates sweat secretion, which can be measured precisely and effortlessly using an EDA sensor on the wrist or feet (Anusha et al., 2020). There is a positive correlation between stress arousal and skin conductance when skin gets saturated (Bari et al., 2021).

Among a limited number of sensors that can be worn outside clinical environments with high accuracy, a non-invasive wearable device is required to measure humans' physiological parameter changes in workplaces such as construction. The data recorded by such sensors can be used to identify stress as the nuances of physiological and biological features correlate with a human's neurological and psychological state (Lean & Shan, 2012). For instance, acute psychological stress triggers an instant physiological reaction (Sevil, Rashid, Hajizadeh, Askari, et al., 2021).

2.4 Stress Detection Methods and Algorithms

Most stress detection techniques rely on classification, regression, and clustering to classify different stress levels experienced by subjects (Gedam & Paul, 2021; Hasan et al., 2022). Algorithms such as Logistic Regression (Rodríguez-Arce et al., 2020; Subhani et al., 2017), Support Vector Machine (SVM) (Betti et al., 2018; Can et al., 2020; Subhani et al., 2017), and Naïve Bayes (Airij et al., 2018; Egilmez et al., 2017; Saeed et al., 2020) are frequently used for classification in stress detection application.

Classification research conducted by Eskandar and Razavi (2020) applied a Long Short-Term Memory (LSTM) algorithm on an open database of multivariant time series collected by Birjandtalab et al. (2016). LSTM algorithm was applied to train and test the method to classify four different neurological states (i.e., four classes). A robust automated pattern recognition method, using a deep learning LSTM, was used to identify states by identifying associated patterns. This study considers the same number of instances for each class in the preprocessing stage. Failing to follow this requirement leads to having more data points associated with one class and hence biases the method to predict more of that class, failing to perform adequately in a testing stage.

Common classification algorithms fail to classify anomalous points. Anomalies are data points in the dataset that differ from other data points (i.e., non-anomalies) which deviate from the dataset's usual patterns. Anomalies occur rarely and are significantly fewer than non-anomalous points in a dataset (Schneider & Xhafa, 2022). Standard classification algorithms fail to distinguish and classify anomalies since a relatively comparable amount of positive and negative samples in a classification problem doesn't support issues with significantly fewer positive (anomalous) instances than negative (non-anomalous) instances.

Designing a method to identify patterns associated with the anomalies requires knowing the data type and preparing it before processing it to represent accurate information. Since physiological signals contain inherent noises and are continuous readings in time, pre-processing is a critical step in designing a pattern detection system. Physiological signals are time series data type, and multiple time series are referred to as multivariant time series. This terminology is reviewed in more detail.

2.5 Multivariant Time Series Anomaly Detection

The definition of anomalies relates to the domain and the type of information they represent. Anomalies are data points in a dataset that differ from regular instances, signifying infrequent experience in a system (Teng, 2010). Three generally accepted categories of anomalies are point anomalies, context anomalies, and collective anomalies. Point anomalies are abnormal occurrences despite the entire dataset. In other words, point anomalies often represent some extremum. Context anomalies are abnormal instances in the context of meta-information (e.g., time or space) associated with the data points. Collective anomalies are a subset of data points (as a collection)

that differ considerably from the whole data set (in terms of data features), where the individual data points are neither anomalies nor considered abnormal in a contextual sense.

Detecting anomalies in the time series dataset differs from the typical dataset as it involves temporal dimensions (i.e., time). Predicted trends, patterns, frequent spikes, regular drops, or seasonal changes are some of the time-related occurrences that require different anomaly detection techniques from a dataset without time or order (Lin & Su, 2019). A simple dataset of multivariant time series is shown in Fig. 2 and the highlighted red intervals are identified as anomalies. Although data point values do not fall within low-density regions, they don't follow typical seasonal trends. Seasonal trends are the recurrent changes in data that occur after a period of time. For instance, line A in Fig. 2 represents a seasonal trend as it increases in value after a certain period of time. Similarly, in Line B, a big drop follows by a smaller drop creating a seasonal trend, which happens regularly.

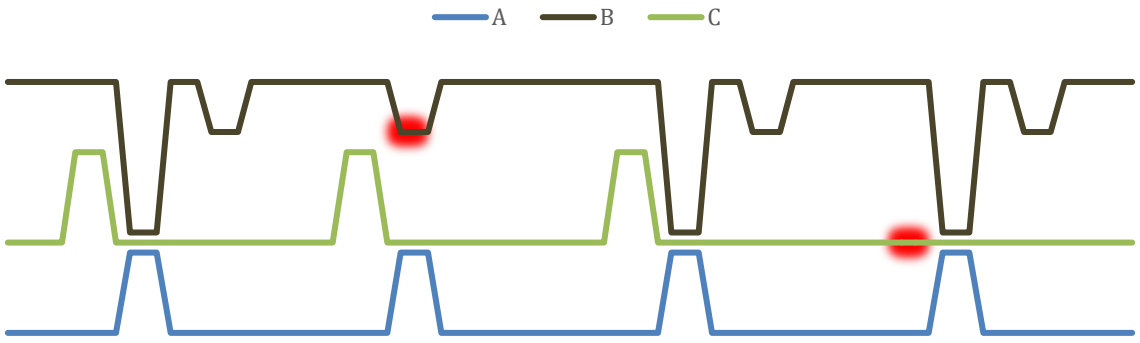

Fig. 2. Highlighted red intervals are identified as examples of contextual anomalies

Unforeseen events and incidences in data can be identified using anomaly detection. Anomaly detection applications are primarily applied to the unlabelled dataset, which requires unsupervised learning techniques (M. Munir et al., 2019). A simple known anomaly detection method that

considers prior information is ARIMA, limited to univariate time series (Piccolo, 1990). Anomalies in multivariate time series are often contextual and generated from a complex interaction of variables. Therefore, identifying anomalies by detecting individual features is erroneous and not reliable.

Deep Learning-based (DL-based) methods have both scalability and long-term temporal dependency advantages over traditional methods (Luo et al., 2021). Among different DL-based methods for anomaly detection, an autoencoder-based method using a simple Recurrent Neural Network (RNN) can be used. Still, it does not demonstrate causal factors contributing to the anomaly. Convolutional Long Short-Term Memory deep learning-based method that considers the causal factors is used for anomaly detection (Deepak et al., 2021).

Time series anomaly detection techniques fall into two paradigms:

(a) Forecasting based and

(b) Reconstruction based methods.

2.5.1 Forecasting based methods

In the forecasting-based method, anomaly detection works based on historical data to predict future data. Several points from the past produce an estimate of the future point and forecasted points in the future will be used to estimate a new point, and it continues. The Auto-Regressive Integrated Moving Average (ARIMA) is a widely used forecasting anomaly detection method that predicts future points using past points and past errors. The ARIMA is used for modelling univariate time series (single variable). LSTM frequently is used for predicting the value in time series based on previous changes. LSTM has feedback connections in addition to the standard feedforward

processing that enables the method to process entire data sequences and make them accessible in time series data. Bayesian Network and Hierarchical Temporal Memory (HTM) also has been used for multivariate time series. These presented forecasting-based anomaly detection methods do not scale well for time series (Razaque et al., 2022).

2.5.2 Reconstruction based methods

In a reconstruction-based method, the method learns how to reproduce the original input by looking at specific latent features. Dobson (2003) described latent variables as "variables that can only be inferred indirectly through a mathematical model from other observable variables." A reconstruction-based method obtains the latent features of an input instance through filters. Filters, also called kernels, can detect spatial patterns and plays an important role. Autoencoder is a neural network architecture to employ the reconstruction-based method, in which every unit in each layer acts as a filter. Autoencoder reconstructs the input instance by first encoding along hidden layers (compressing input data by reducing value dimensions and applying filters) and then decoding (rebuilding the output by increasing value dimensions).

Zhang et al. (2019) compared different anomaly detection techniques fitting multivariant time series and proposed the Multi-Scale Convolutional Recurrent Encoder-Decoder (MSCRED). The proposed method consists of four Convolutional Long-Short Term Memory (ConvLSTM) layers with attention module. The method was compared to eight baseline methods in four different categories:

(1) Classification models to challenge the model's decision functions:

- One-Class SVM model (OC-SVM)

(2) Density estimation models to detect models' data density impact for the purpose of outlier detection:

- Gaussian Mixture Model (DAGMM)

(3) Prediction models to test models' temporal dependencies over training data and its' strength in predicting the value of test data:

- History Average (HA)
- Auto-Regression Moving Average (ARMA)
- LSTM encoder-decoder (LSTM-ED)

(4) MSCRED variants to justify the effectiveness of each component.

- $\mathrm{CNN}^{\mathrm{ED}(4)}_{\mathrm{ConvLSTM}}$ with attention module and 4th ConvLSTM layer
- $\mathrm{CNN}^{\mathrm{ED}(3,4)}_{\mathrm{ConvLSTM}}$ with attention module and 3rd and 4th ConvLSTM layers
- $\mathrm{CNN}^{\mathrm{ED}}_{\mathrm{ConvLSTM}}$ without attention module

The proposed method, MSCRED, is $\mathrm{CNN}^{\mathrm{ED}}_{\mathrm{ConvLSTM}}$ with attention module, outperformed baseline methods. The MSCRED captures spatial and temporal dimensions to understand the system's behaviour (Zhang et al., 2019) .

This research takes a deep learning autoencoder equipped with Long Short-Term Memory (LSTM) plus an attention module as the preferred method for multivariate anomaly detection problems. Autoencoders are deep learning networks where Convolutional Neural Network (CNN) reconstructs the same input following the same pattern as explained above. For the chosen method, spatial dimensions and inter-feature relationships are captured through correlation matrices as input. The selected method labels an input as an anomaly if the reconstruction error is high (Z.

Cheng et al., 2021), in which reconstruction error refers to a difference between the original input and reconstructed output. This method is deep learning as it holds several hidden layers.

2.6 Cyber-Physical Systems for Construction Safety

Human-in-The-Loop Cyber-Physical System for Workplace Safety Industry 4.0 has led to rapid technological and industrial changes in the 21st century due to rising interconnectivity and automation. The construction industry, as one of the main economic sectors, has benefited from automation and integrated systems. Some of the applications are; Virtual Reality (VR) and Augmented Reality (AR) applications for civil infrastructural projects (Behzadan et al., 2015; Xie et al., 2022), automation in controlling safety measures (J. Wang & Razavi, 2019); visualization of a construction zone (Guo et al., 2017); identifying the ongoing activities and progress in working areas (K. Liu & Golparvar-Fard, 2015); preventing collision and contact injuries by warning the construction workers and heavy equipment operators when hazardous item proximity is detected (Genders et al., 2015); digital twin (Ruikar et al., 2021) and improving the delivery process in building and infrastructure development (Anumba et al., 2010; Hegazy et al., 2014).

Technology and engineering advances have led to robust connections between the physical world and cyberspace, assisting civil engineering in various areas, from smart design to automatic heavy equipment. Common examples of these systems are Cyber-Physical Systems (CPS), in which the physical setting is tightly connected with computation sections. CPSs are aimed to observe and control the physical environment intelligently. These systems interact with their operating environment mainly for the purpose of measuring and sensing the physical features, process the information, and then take actions to reach the desired outcome. It uses sensors and actuators to

monitor and control physical phenomena and create monitored, maintained, and adaptable environments (Nunes et al., 2018).

CPS reforms the way humans interact with the physical world through linking cyberspace and the physical world. In construction setting, to facilitate the flow of information in the system, the cyberspace can connect to the physical space through the integration of virtual design and physical actuators. Humans or users are essential to such systems (Munir et al., 2013). Notably, in construction sites, where heavy machines and workers are moving, working, and interacting in a complex, dynamic, and hazardous environment, human interaction with the environment is a primary concern. Including a human entity in a cyber-physical system can result in more resilient information acquisition (Cárdenas et al., 2009; Cooke & Chong, 2017; Leitão et al., 2016). A resilient flow of information is valued due to the lack of awareness and the excess mental workload in a construction zone (Cooke & Chong, 2017; Griffor et al., 2017; C. Zhang et al., 2017).

A CPS that considers humans in the system is called Human-in-The-Loop Cyber-Physical System (HiLCPS or HiLCPS). The HiLCPS has remarkably enhanced performance and accuracy due to adding humans to the main control loop (Nunes et al., 2018). It incorporates human intentions, behaviour, presence, psychological state, emotions, and actions, which help better understand the context and determine more appropriate measures. By studying the context, systems' goals, and the human-related factors affecting the system's performance, we gather information to define a role of a human entity in the HiLCPS system.

Refining construction safety has been a critical concern due to the industry's high rate of injuries and fatalities and the environment's complexity and unpredictable nature (van der Molen et al., 2005). Accidents mainly happen in construction due to unsafe acts and conditions, for which

unsafe behaviour and error by humans play a critical contributing factor (Chua & Goh, 2004; HEINRICH, 1950; Reason, 1991).

Multiple frameworks for HiLCPS are developed (Griffor et al., 2017; S. Munir et al., 2013; Schirner et al., 2013). In the framework by Schirner et al. (2013), a human role was studied in relationship with the embedded system (cyberspace) and the physical world, in which sensors collect information on the human state and intentions from brain or body. Collected information is transferred to the cyber system to be processed and transformed into a readable version by the physical world. In the end, appropriate action is performed using actuators to achieve desired state. In another taxonomy by Munir et al. (2013), the human connection with the HiLCPSs are classified into three forms: systems controlled by humans, systems passively monitor humans, and a combination of the two. Nunes et al. (2018) proposed a taxonomy in which human was considered in three stages: (1) data acquisition, (2) state inference, and (3) actuation. Each stage plays a crucial role in collecting data, delivering status assessments, or operating in the actuation stage.

The latest taxonomy by Nunes et al. (2018) has both simplicity and inclusion and benefits from an added feedback control. Considering feedback in a loop can enable the system to perform with higher efficiency and flexibility (Griffor et al., 2017; Guan et al., 2016; S. Munir et al., 2013; C. Zhang et al., 2017).

Humans can be placed in any stage depending on the application and the context. Knowledge regarding human factors and parameters in favour of the system goals are crucial for incorporating human in each application. Detecting and understanding human contributing factors causing a dangerous action is essential for design and development of a HiLCPS safety application. The following sections describe the human role in HiLCPS, as shown in Fig. 3.

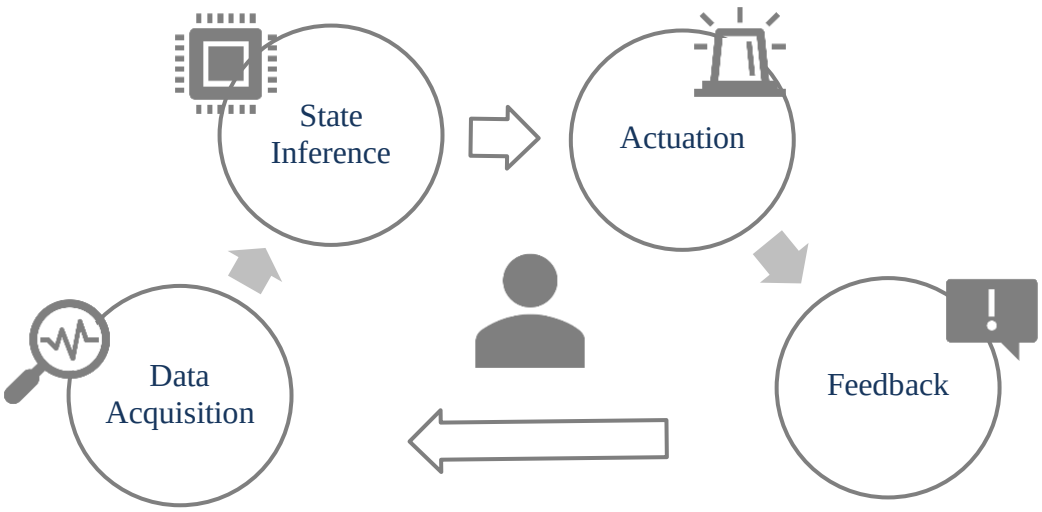

Fig. 3 Different human roles in HiLCPS for Workplace Safety

2.6.1 Data Acquisition

Humans add to data acquisition through either sensor connected to a human or providing information directly to the system. Examples of human-generated data could be social media activities, contact lists, or setting preferences by human. Collected information can be transferred actively or passively. Data acquisition consists of collecting data on physical phenomena such as body temperature and blood pressure or nonphysical like social media posts or human. Such information enables the system to detect patterns related to human physiological features, human psychological states, and external conditions. Selecting sensors and data gathering methodology is essential depending on the application. For instance, wearable sensors are recommended to avoid disruptions to the work and workers' mobility or focus. Also, non-disruptive ways of transferring collected data such as Wireless Sensor Networks (WSN) with data streaming services need to be considered (Wood & Stankovic JA, 2008).

Collecting real-time information in an undisruptive manner has become more affordable over the past decade. Human physiological signs such as heart rate, oxygen level, and sleep schedule can be recorded without disruption through wearable sensors (e.g., health gadgets, smart watches, smart rings, and smart fabrics). Moreover, information such as electrophysiological signals[1] can be measured using body and brain sensors (Schirner et al., 2013).

Smartphones are great data collection devices that enable passive and accessible data collection. They are capable of sensing human activities and processing the information. Smartphones at the edge of the network referred to as edge computing, in which analyze the recorded data at the edge rather than transferring data to the core network or cloud data center for processing (Bonomi et al., 2012; Cisco Systems, 2016).

2.6.2 State Inference

The state inference or data processing stage assists in inferring human intent from the collected information. This process can detect physiological or psychological states such as stress or overload through asking human for processed data or using mathematical or machine learning models and infer the states. So, the role of human in the state inference could be offering direct information or expert knowledge (Nunes et al., 2018).

[1] electrophysiological signals are electrical activity in different parts of the human body, such as electroencephalography (EEG), electrocardiography (ECG), and electromyography (EMG)

2.6.3 Actuation

Actuation can refer to a notification on a cellphone to alert a user. A human's contribution in the actuation step is when there is a need for human action than robotic actuators. Humans and machines can collaborate in actuation, and their functions can be supplementary. Depending on the identified state, actuation changes the physical environment to reach the desired goals. Consider a simple application like a slip or fall detector that records a worker's physical information using an accelerometer. This system can translate a rapid change of acceleration value in the vertical axis to a fall. Instant alarms and notifications may be sent to the safety controller and medical emergency services.

2.6.4 Feedback

Feedback controls some parts of the system, such as changing preferences, which is an example of incorporating humans into feedback control. Humans' contribution to this process will result in a more reliable and robust system (S. Munir et al., 2013). It's common for a system to face some changes or challenges in its life cycle that might contradict its performance. Feedback from a user can improve and direct the system toward the desired outcomes. Systems, in addition to the self-adopting process, can learn from human through observation (Z. Liu et al., 2011).

Adopting HiLCPSs in workplace safety applications can be summarized in sensing and monitoring physiological factors to identify unsafe human actions and conditions and assist the user with information to improve safety (Gualdi et al., 2009). In terms of a physiological aspect, bending posture, fatigue, and stress contribute to unsafe behaviour. In HiLCPS application, the associated data needs to be collected (e.g., workers' tri-axial acceleration for posture analysis detection) and

data can be inferred to understand the system state (e.g., calculating bending angle or observing fatigue symptoms). Then, suitable action can be taken to improve workplace safety.

2.7 Research Gaps

Based on what has been discussed so far through reviewing the literature, the existing stress detection studies are mainly focused on and differ from each other in the following aspects:

(1) Data collection method (e.g., wearable sensors, pupil tracking, questionnaire);

(2) Selected features and variety of physiological signals;

(3) The stress-inducing approach and design of the field experiment (e.g., data collection location and designed activities);

(4) Learning algorithms;

The review of physiological data associated with stress and overload in section 2.3 revealed that the current stress detection methods based on physiological signals could benefit from further situational or contextual input data, such as the subjects' basic and demographic information. Moreover, there are inadequate stress-inducing activities in conducted field experiments in stress detection studies, which is discussed in more detail in Chapter 4 section 4.2. A critical finding by studying the background revealed the lack of anomaly detection algorithms in stress detection, while moments of unsafe stress states can be considered anomalies. This research differs from the existing literature in (1) input features, (2) various stress-inducing activities in the field experiment, and (3) the use of multivariate anomaly detection algorithms. Moreover, it aims to train a stress detection algorithm that (4) provides an interpretable method for easier comprehension. Deploying such a method requires studying current integration systems, which is presented in section 2.6.

Chapter 3 Methodology

3.1 Chapter Overview

Stress as an essential influencing factor to unsafe behaviour in the workplace motivated this study on intelligent stress detection methods to detect human abnormal stress states for occupational safety. Stress is detected through a combination of contextual data and variations in physiological features such as heartbeats, blood volume, and skin temperature. This chapter presents the research methodology in section 3.2.

3.2 Research Methodology

The following methodology framework is used to achieve the objectives presented in section 1.3. This section offers the general roadmap to research methodology, shown in Fig. 4 along with details and defined activities. The framework, presented in Fig. 4, has six broad sections, namely;

(1) Preliminary Stage,

(2) Data Collection,

(3) Data Pre-Processing,

(4) Stress Anomaly Detection,

(5) Validation, and

(6) Conclusion.

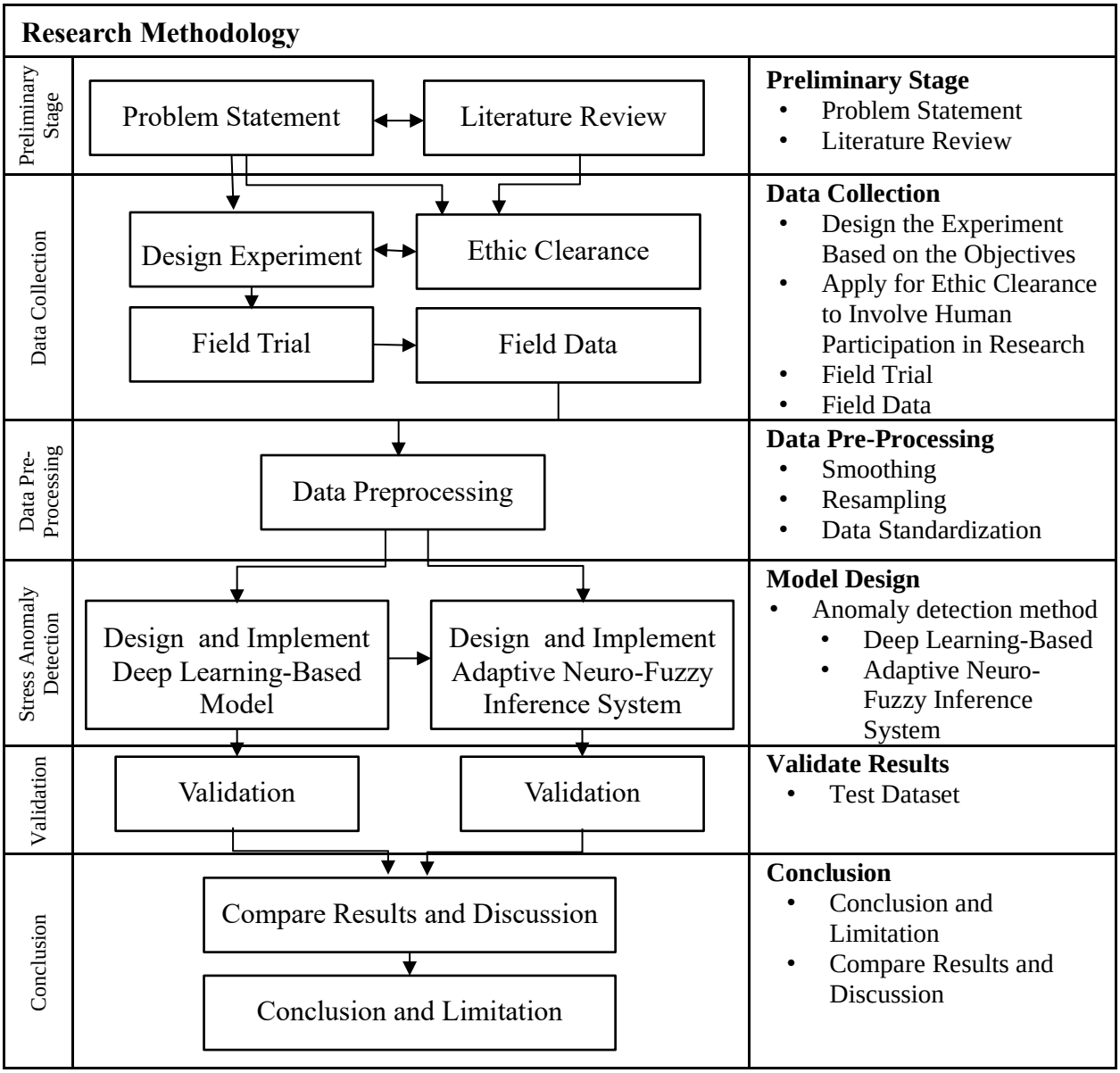

Fig. 4 Research Methodology

3.2.1 Preliminary Stage

The preliminary stage of the research started with a literature review for the problem statement and identifying research gaps to define the scope and objectives of this study. An initial literature review regarding unsafe human behaviour and accidents' root causes led this research to focus on

human stress and overload. A wide range of literature reviews on physiological data associated with stress, data collection method, as well as stress detection methods and algorithms followed. An overview of a cyber-physical system for potential use in construction safety applications is presented, followed by the role of humans in the loop of a cyber-physical system.

3.2.2 Data Collection

A field experiment is designed and proceeded for ethical clearance to involve human participation in the research stated. Due to the COVID-19 pandemic, the designed experiment was revised, and the ethics clearance was approved for conducting research when the field experiment location (i.e., McMaster Campus) is outside the Covid-19 municipal red zone. Following the Covid regulations and gathering participants' consent, the experiment was conducted safely with 18 participants in 18 appointments. Participants were recruited and compensated following a McMaster Research Ethics Board (MREB) procedure.

Three sets of data were collected during each field trial: (1) multivariate physiological signals. The participant’s physiological features were collected while performing specific activities that triggered mental stress and overload. A wristband biosensor was used for the field experiment to collect real-time electrodermal activities, skin temperature, blood volume pressure, 3-axis accelerometer, and heart rate while participants followed planned activities. (2) basic-level data such as demographic information, height, weight, hours of sleep, and participants’ mood; and (3) Situation-level data such as activity type and the participant's reaction time at specific time intervals during the experiment.

3.2.3 Data Pre-Processing

Data preprocessing is essential to fuse the above-noted three levels of data. As explained, real-time human physiological signals are collected using wearable sensors with different frequencies that require noise removal and alignment. Data smoothing for noise cancellation applied, alignment including up-sampling and down-sampling, and standardization.

3.2.4 Stress Anomaly Detection

A wide range of literature on stress detection methods and algorithms motivated the adoption of an unsupervised deep learning method. To detect occupational stress from physiological signals, there is no formulated equation or algorithm to calculate, and there are countless contributing factors affecting stress levels. So, the nature of the problem promotes deep learning-based applications.

3.2.4.1 Deep Learning-Based Anomaly Detection

Unsupervised deep learning-based, inspired by the anomaly detection methods presented by Zhang et al. (2019) adopted that looks into a correlation matrix of 14 collected features. Using Long Short-Term Memory (LSTM) helps the method to keep track of fluctuations over a few seconds. The method learns countless patterns by going through 500k matrices of size 14×14. Using correlation matrices and looking into the correlation between features equips the method with an interpretation tool to demonstrate the contributing features.

3.2.4.2 Adaptive Neuro-Fuzzy Inference System

This method presents a proof of concept for the potential use of Fuzzy Inference Systems (FIS) to detect the degree of anomaly. Adaptive Neuro-Fuzzy Inference Systems (ANFIS), as a form of

FIS-based, is selected due to the absence of reasoning about the inherently vague concept of abnormal psychological states and the shortcomings of the first method. The limitations of the first method are a need for thousands of training samples, the high cost of training and data collection, over-reliance on uncertain absolute truth, and setting a threshold for detecting anomalies. The shortcomings motivated this research to investigate further. Fuzzy logic promotes solving problems with an imprecise spectrum of data that enables finding an array of accurate conclusions. Chapter 6 presents the proposed Adaptive Neuro-Fuzzy Inference Systems (ANFIS) and its architecture. In the second method, the input, output, and rules are modelled as neurons, in which the values of fuzzy sets are transferred as weights. Section 6.3 describes the basics of the method and the formulas, and section 6.4 describes the detail of the proposed ANFIS for the anomaly detection problem. Then, section 6.5 describes experimental results and discussion, and the last section presents the conclusion.

3.2.5 Evaluation

The proposed methods in Chapter 5 and Chapter 6 are evaluated in sections 5.5 and 6.5 against two separate test datasets. The labels (i.e., absolute truth) for the first test dataset are tagged on the expert's opinion by studying the raw collected time series. The second test dataset was arranged without following the original sequence of activities in order to prove the method's independence from the sequence of activities in the designed experiment. While recording data, the times of intentionally added extra mental stress/overload were logged as labels. Precision, recall, F1-score metrics for each class as well as accuracy, macro averaged, and weighted average are calculated to validate the performance of the methods.

3.2.6 Conclusion

The final section includes a methods comparison with regard to efficiency, training cost, and performance metrics. It also presents research contributions as well as limitations and future works for the field experiment, unsupervised deep learning, and adaptive neuro-fuzzy inference system.

Chapter 4 Data Collection and Pre-Processing

4.1 Chapter Overview

This chapter is allocated to all related information regarding the field experiment and collected data. As explained in section 1.3, we aim to design and implement a field experiment to collect sample physiological data while inducing stress. The field trial design, equipment, and activities that triggered mental stress and overload are expanded in section 4.2. The participant's physiological features were collected during the trial, presented in section 4.3, which includes three sets of data. Section 4.3.3 is assigned to signal-level data that is collected using a biosensor wristband, and section 4.4 expands the preprocessing procedure, such as smoothing, resampling, and standardization. Ethical clearance must be achieved due to human involvement in the field experiment, and section 4.5 is allocated to participants' ethics, health, and safety. The ethics clearance is approved for the following proposed experiment under MREB #5015.

4.2 Experiment Design

The experiment is designed with a sequence of activities to expose the participant to stress and overload. In the planned experiment, it is anticipated that each participant will experience mental stress and overload while following the proposed stress-triggering activities. Throughout the experiment, real-time physiological features were constantly recorded using a wearable biosensor

to monitor the physiological reaction. Similar studies have adopted chosen activities as mental stress-inducing activities that reduce cognitive performance and induce stress (Bali & Jaggi, 2015; Brisswalter et al., 2002; Calibo et al., 2013; Rao et al., 2017; Renaud & Blondin, 1997). Stressors are events or conditions that have the potential to stimulate stress responses. There are different categories of stressors: physical, environmental, mental or task-related, social, psychological/ emotional, chronic, and traumatic (Cooper, 2005). In this study, a combination of physical and mental activities has been adopted as the ultimate objective is workplace safety and involves both stressor types.

Due to their effectiveness, the Stroop Colour Word Test and mental arithmetic tasks were identified as the most used stressors in inducing mental or task-related stress (Giannakakis et al., 2022). The Stroop colour naming test uses visual and written stimulation that has been widely used (Hjortskov et al., 2004; Hou et al., 2016; Salahuddin et al., 2007). Fig. 5 presents the Stroop test interface. Human physiological reaction upon solving the Stroop test has been validated (Tulen et al., 1989). Similarly, mental arithmetic exposure has been validated as a physiological stimulus that induces stress (Lackner et al., 2011).

Among mental or tasks related stressors, arithmetic problems (e.g., counting backward by 7), video games (e.g., speed race game), and solving puzzles are frequently used to induce stress. Exceptionally, video games, such as a speed race game with a moving level of difficulty that requires high cognition performance, are efficient in generating stress (Karthikeyan et al., 2011; Rani et al., 2002). Fig. 6 presents an interface for the speed race test. The user needs to dodge moving and static obstacles in a race.

Physical activity and keeping balance have also been adopted as they elicit physiological stress reactions and have been proven to be stress stimuli (Kerr et al., 1985; Sevil, Rashid, Hajizadeh, Park, et al., 2021). Demographic information determines the intensity of the physiological reaction (Giannakakis et al., 2022), in which related demographic information, planned to be recorded through a questionnaire, is explained in section 4.3.1. There are proven differences between physical and psychological stress and how they affect the body (Ponce et al., 2019).

Fig. 6 Stroop colour word test interface

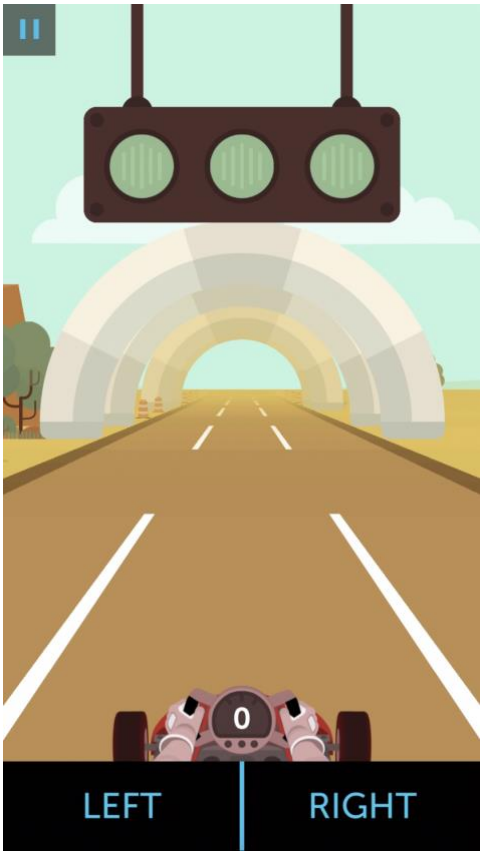

Fig. 6 Speed race game interface

This experiment's activities are designed and undertaken based on the previously reviewed research. At the beginning of each experiment, the participant was asked to relax for 5 minutes

(warming up period). The relaxation period was then followed by different activities in the following order:

(1) **One-foot balance**, in which participants were asked to stand on one foot at a time and keep the balance for two minutes.

(2) **Walking**, in which participants follow a specific path while walking at an average pace for three minutes. The route was marked by tape on the ground to guide the subject in the desired direction.

(3) **Speed race game**, performed on a digital device; participants must skip obstacles in a race through the desert with moving tumbleweeds, other drivers, and other barriers.

(4) **Stroop colour word**, performed on a digital device, measures participant's flexibility. Two colour words appear horizontally on the screen in different colour fonts (Fig. 5). Participants were asked if the meaning of the upper colour word matches the colour font of the lower colour word.

(5) **Counting backward**, in which participants were asked to count backward by seven from 2864 (e.g., 2864, 2857, 2850).

Participants were asked to perform under time pressure for the speed race game, Stroop colour word, and counting backward. A five-minute relaxation session was given to reset the psychological state to a relaxed state between each activity. The 5 minutes of relaxing time between different tasks ensure the mental state is back to baseline, and the order of tasks doesn't impact the result. Moreover, certify that the order of activities is independent of the results obtained. Fig. 7 demonstrates the timeline for the proposed experiment.

Fig. 7. The timeline of the experiment

After each activity, the Ruler test was conducted, which provided us with the reaction time. More detail on the time and sequence of the ruler test can be found in Fig. 7. In the ruler test, the test-supervisor drops a ruler vertically between the participant's thumb and index finger and the participant catches the ruler between their fingers. The test supervisor records the measurement on the ruler where the participant's fingers are. This test is used to assess the participant's reaction time. The ruler test was repeated three times (for every reading) to gain the average reaction time. Fig. 7 illustrates the timeline of the conducted experiment. Individuals were asked to wear long pants and boots in addition to personal protective equipment (PPE), a safety hat, a vest, and safety shoes. Instructions for each activity were given immediately before the activity to increase mental

stress and overload during the experiments. The test supervisor observed participants during activities and stayed out of sight for the relaxation sessions to take away the mental stress of being watched.

4.3 Data Collection

Three types of data collected during the proposed experiment are discussed below. Data collection methodology, i.e. questionnaires, labelling, and clinical wearable device, have been chosen for basic, situational, and signal-level data, respectively. The following three subsections present more information and examples for each type.

4.3.1 Basic-Level Data

At the beginning of the experiment, basic-level data is collected using a questionnaire to gather participants' demographic information. Basic-level data contain participants' gender, age, height, weight, mood, and hours of sleep. The following table represents the coding and value assignment for the basic-level data for some of the participants of the field experiment.

Table 1: Portion of basic-level data

ID	Gender	Age	Height	Weight	Feel	Sleep
S1	F	29	165	62	3	2
S2	F	30	165	57	3	4
S3	M	38	187	98	3	1
…	…	…	…	…	…	…
S18	…	…	…	…	...	…

4.3.2 Situation-Level Data

The second type of data is the records of activities and the exact time of changes, also called situation-level data. This category presents the performed activity and reaction time throughout the experiment and shows the type of work a participant performs, which varies over time as the activity changes. Internal real-time clock with high accuracy of 5ppm[2] in the adopted wearable wristband enables precise time reference. The following section presents more information on the wristband.

Table 2: Sample of situation-level data

Time Index Tag Occurrence	1609977632	1609977763	1609977825	1609977946	1609978262	1609978453	1609978755	1609978946	1609979247	1609979437	1609979742	1609979957
Activity Label	Relax	One foot Balance	Relax	One foot Balance	Relax	Walking	Relax	Speed	Relax	Stroop	Relax	Counting
Ruler test (cm)	30			42		25		24		22		15

[2] Parts per million (ppm), is defined to denote the frequency difference of two clocks. It's a way to compare accuracies.

4.3.3 Signal-Level Data

Signal-level data, as the third type, was collected using a wearable biosensor. Despite the ability of clinical / in-lab devices that enable a broad range of information collection (e.g., EEG headset, cortisol secretion level), a non-invasive wearable wristband is selected for this research so that the method can be transferable to workplace environments. Furthermore, due to the nature of the work and the necessity for real-time reading of physiological features, a wearable sensor with a Bluetooth streaming mode was selected that also enables easy access to raw data with a secure cloud platform. Clinical quality precession is another feature of the chosen wristband that precisely measures the sympathetic nervous system and heart rate simultaneously.

(a) (b)

Fig. 8 Empatica E4 wristband (a) front and (b) inside view

Error! Reference source not found. shows the selected wearable biosensor from the front and inside angles. This wristband collects five physiological features: blood volume pulse (BVP),

photoplethysmography (PPG), skin temperature (ST), 3-axis accelerometer (ACC), and heart rate (HR). The above-noted physiological features are collected using the following embedded sensors:

1. PPG sensor to record blood volume pulse, which is used as an indicator of heart rate variability (HRV) and cardiac changes such as inter-beat-interval[3] (IBI);
2. Galvanic skin response (GSR) sensor to measure the changes in electrical properties of the skin, which is an indicator of sweating rate;
3. Infrared Thermopile to read peripheral skin temperature, which is an indicator of thermoregulation, and
4. ACC to capture motion-based activities as an indicator of physical movement.

The raw multivariate physiological data, recorded by wearable wristband, is presented in Fig. 9. The vertical axis indicates the timeline of the experiment, and horizontal red lines are tagged time during the experiment at which the activity is changed. The horizontal axis shows different physiological measurements.

4.4 Data Pre-Processing

Pre-Processing and integrating wireless sensor data is a primary step toward having an efficient, robust, stable, quality-assured model tested with an actual data set. Therefore, the acquired signal-level data is preprocessed to reduce noise, align frequencies, and scale. More details of the applied

[3] The IBI, inter-beat-interval, is calculated by detecting the BVP peaks and computing the lengths of the intervals between adjacent peaks

pre-processing methods using Python language programming can be found in Pre-Processing Jupyter Notebook (Eskandar, 2022c). The following sections briefly discuss these processes.

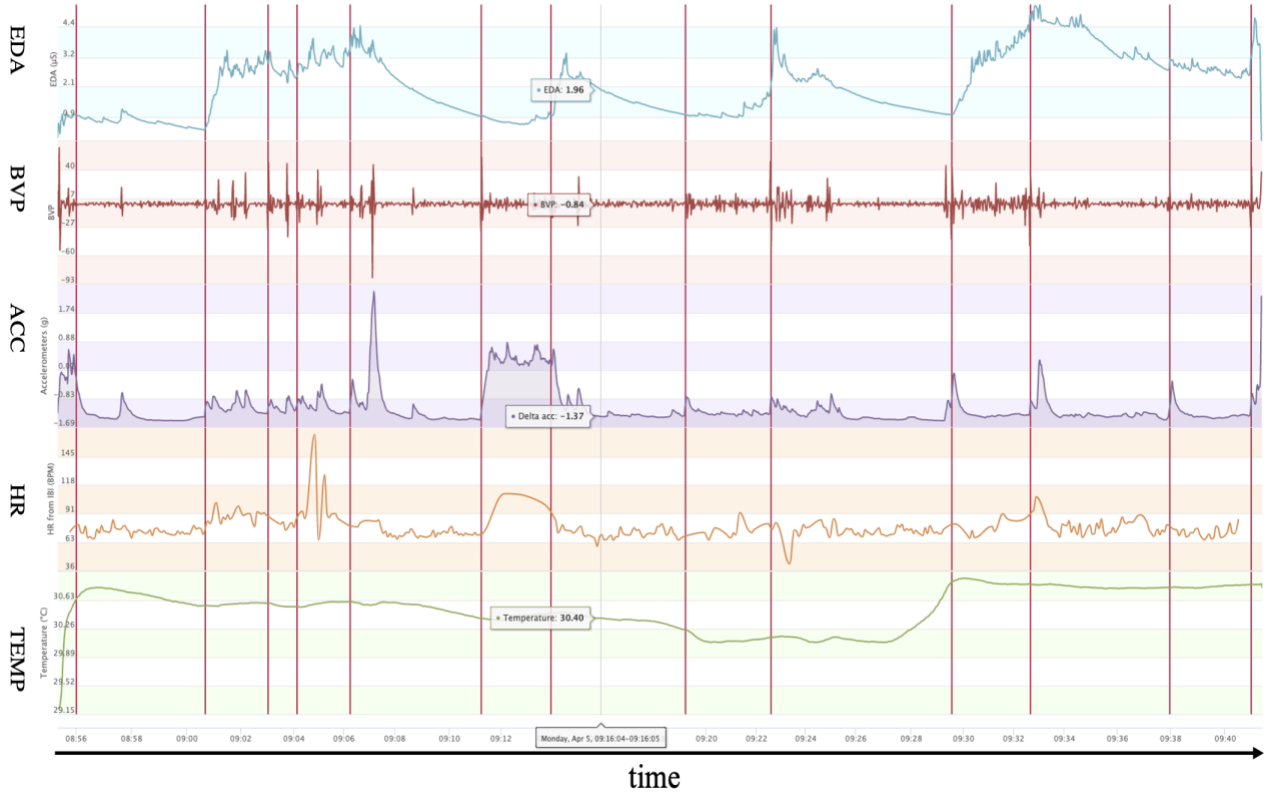

Fig. 9. Sample of signal-level data

4.4.1 Smoothing for Noise Reduction

If data is measured by sensors or is generated from a physical phenomenon, it inherently contains noise and requires noise reduction. An individual's biometric signals captured through wearable sensors are susceptible to external (i.e., environment artifacts) and internal (i.e., physiological artifacts) noises. There are many ways to reduce wearable sensor noises without removing important information. Simple yet effective methods such as normalization (shifting and rescaling

data points to a range of 0 to 1) are applied. Normalizing the input signal-level data improves robustness and efficiency by transforming raw recording into a uniform pattern.

The Locally Weighted Scatterplot Smoothing (LOWESS) function (Tibshirani & Wang, 2008) is a widely used method that recovers a signal from noise through localized measures by looking at each data point and calculating the smooth value using its neighbouring samples (d samples in the D set). The LOWESS method used in this study is similar to K-Nearest Neighbour (KNN) algorithm (Fix & Hodges, 1989), in which window size k determines the neighbouring sample. A set of weights is applied to adjacent samples to reflect the importance and distance to the data point. The larger k (or window size) in localized measures results in lowers variance and higher bias. The optimum size should be identified to remove and flatten external noises and not remove information. Accurate smoothing assists in better understanding the signal by leaving behind the noises and discovering patterns.

4.4.2 Data Resampling

The collected multivariate time series have different frequencies besides the extrinsic signal artifacts that require preprocessing. By up-sampling the data with smaller frequency and down-sampling data with larger frequency, raw recorded multivariate time series are resampled to 4 Hz. IBI and Heart rate samples with 1 Hz frequency are up-sampled using interpolation to 4 Hz frequency. EDA (8 Hz frequency), PPG (64 Hz frequency), and accelerometer (32 Hz frequency) readings are down-sampled with a median of 4 Hz frequency. As such, all signals are aligned for each point in time (at each timestep). Resampling might differ based on the frequency of signal-level data and chosen frequency for aligning information.

4.4.3 Data Standardization

Input features have different units (i.e., g, μs,°C, beats per minute, kg, cm), which requires feature scaling in the pre-processing step. Feature scaling is crucial for distant-based algorithms (e.g., KNN, K-mean, SVM) and gradient descent-based algorithms (e.g., logistic regression, neural network). Several scalers were applied to the recorded data, and a portion of the scaled information is presented in graphs to visualize the scaling method outcomes. Methods used for feature scaling are Standard Scaler, Min-Max Scaler, Robust Scaler, and Normalizer.

Standard scaler can be applied over a normally distributed dataset using Eq. (1), in which $\mathrm{mean}(x)$ is the distribution mean and $\mathrm{stdev}(x)$ is distribution standard deviation. Fig. 10 shows scaled values of physiological time series using the Standard Scaler formula.

$$x_i^{\mathrm{Scaled}} = \frac{x_i - \mathrm{mean}(x)}{\mathrm{stdev}(x)} \qquad \text{Eq. (1)}$$

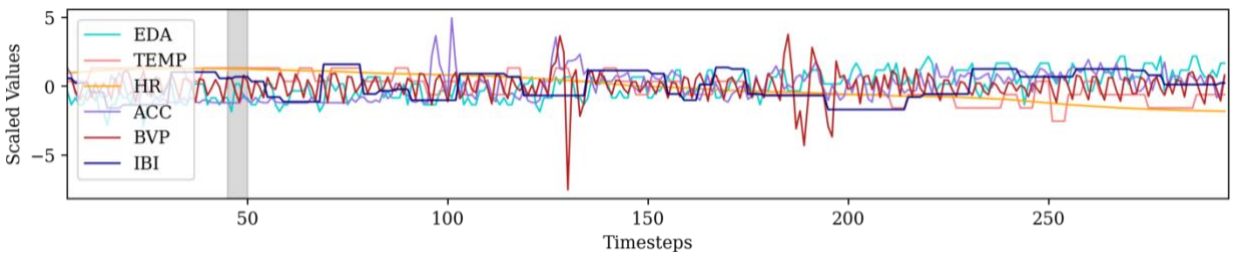

Fig. 10. Physiological data scaled value using Standard Scaler

The MinMax Scaler, a famous scaling algorithm, shrinks the range to 0 and 1 or -1 to 1 (in the case of existing negative values). This scaler is sensitive to outliers, and it is best if the standard deviation is minimal or the data is not normally distributed. Eq. (2) demonstrates the MinMax Scaler formula and scaled physiological time series using this scaler are presented in Fig. 11.

$$x_i^{Scaled} = \frac{x_i - min(x)}{max(x) - min(x)}$$ Eq. (2)

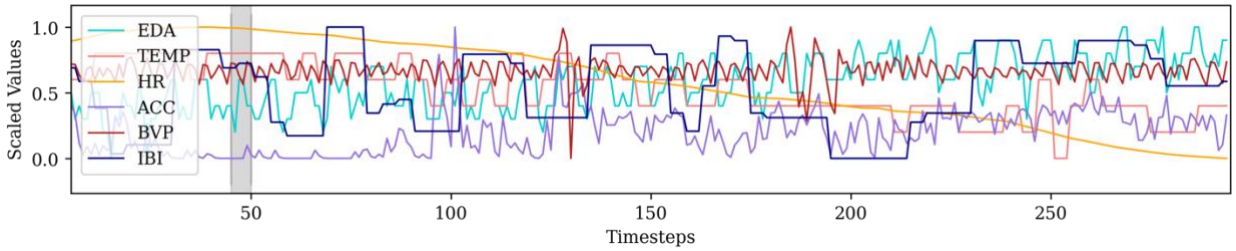

Fig. 11. Physiological data scaled value using MinMax Scaler

Robust Scaler is similar to the MinMaxScaler, but the interquartile range is replaced with the min-max range, which makes it robust to outliers. Eq. (3) presents the Robust Scaler formula, in which, $Q_1(x)$ is first quartile and $Q_3(x)$ is the third quartile. Fig. 12 presents physiological data scaled time series using the Robust Scaler.

$$x_i^{Scaled} = \frac{x_i - Q_1(x)}{Q_3(x) - Q_1(x)}$$ Eq. (3)

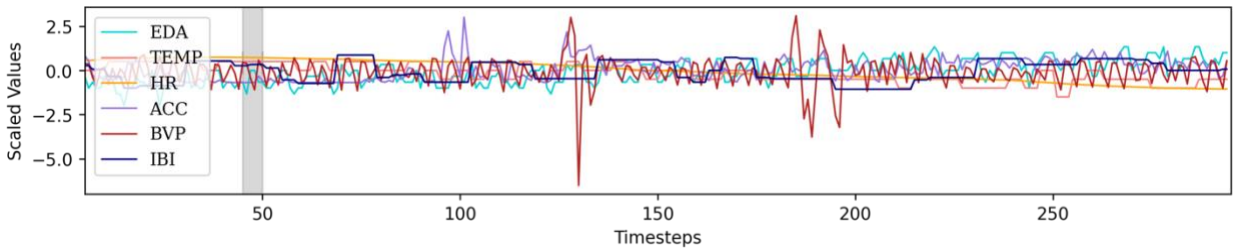

Fig. 12. Physiological data scaled value using Robust Scaler

The Normalizer scales each value by dividing it by its magnitude in an n-dimensional space. Assume x, y and z dimensions of a feature. The scaled value for x would be calculated in Eq. (4). Fig. 13 presents physiological data scaled time series using Normalizer.

$$x_i^{Normalized} = \frac{x_i}{\sqrt{(x_i^2+y_i^2+z_i^2)}} \qquad \text{Eq. (4)}$$

Robust Scaler has been chosen among different standardization methods since a hyperparameter tuning stage resulted in a lower error and higher accuracy, which will be discussed in section 5.5.

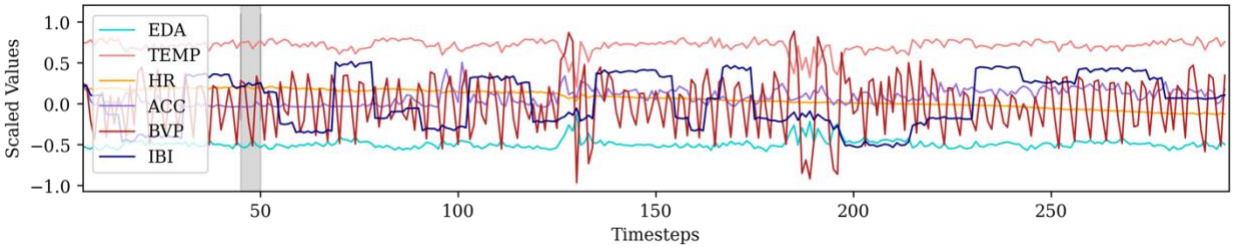

Fig. 13. Physiological data scaled value using Normalizer

4.5 Ethics, Health and Safety Measures of the Data Collection Experiment

An experiment brochure and a poster were prepared and posted to recruit participants on online platforms associated with McMaster University, e.g., McMaster Facebook groups and Instagram pages. Individuals interested in this experiment contacted the researcher via email and had their questions and concerns answered using the same communication channel. Due to the pandemic restrictions, the experiment took place at McMaster University's campus area under extensive cautionary measures to protect the health and safety of all participants and the test supervisor. Informed consent was obtained from all participants before experiments, including a Letter of Intent (LOI) and McMaster Covid-19 Letter of Intent (Covid LOI) for precaution. Section 4.5.1 summarize related details of COVID-19 risks and procedures for in-person research at McMaster University, and section 4.5.2 presents the LOI.

4.5.1 COVID-19 risks and procedures

The experiments occurred at McMaster's main campus, in the open area on the northwest side of the University Hall. All on-campus guidelines, such as maintaining a distance of 2m and wearing a face-covering mask, were considered in the experiment's design. All devices and tools were sanitized with extra caution after each experiment, and the required tools and instruments were placed on a desk for picking up by the participant so that no close contact was needed for the whole process. Before the session began, the responsible investigator disinfected all touching surfaces with disinfectants from the list of hard-surface disinfectants with evidence against COVID-19 (Government of Canada, 2020). The wristband and earbuds are disinfected by allergy-tested disinfected from the same list on Canada's government website.

Only one experiment was organized per day, in which every experiment required a single subject at a time. A social distance of more than 2 meters was kept between a subject and the investigator. Before the session started, individuals were asked to fill out the demographic information forms using paper and a pencil placed on a desk. Individuals were asked to wear Personal Protective Equipment (PPE) (listed in the previous section under test instruments) and a wristband sensor, which is easy to wear, and no close contact was needed.

Participants were asked to wear earbud headphones under hearing protection to receive guidance through the process from a distance for communicating purposes. By communicating through headphones, participants stay at a safe distance from the research investigator while receiving instructions verbally. Participants were allowed to remove the mask and breathe normally in an open environment far from people.

A letter of intent is written in plain language. Using simple and direct language makes it easier for the readers to comprehend. Everything was explained step-by-step, and participants were asked to sign this form right before the experiment.

4.5.2 Letter of Intent (LOI)

Research Sponsor: NSERC Discovery Accelerator Supplement Grant No: RGPAS-2018-522618, and in part by the NSERC Discovery Grant No: RGPIN-2018-05434

Purpose of the Study: In search of representative factors influencing construction safety, stress and overload were among the key reasons attributed to accidents caused by human error. This study includes people aged 18-55 without mobility issues or hearing impairment. This research focuses on identifying and detecting an unsafe state of the human body and mind, which causes insufficient situation awareness and decreased alertness. Individual physiological features will be measured through a wristband that collects real-time electrodermal activities, skin temperature, blood volume pressure, a 3-axis accelerometer, and heart rate. Participants should follow activities that trigger stress and overload, such as standing on one foot, walking on a specific path, playing a speed test game, a Stroop colour-word test game, and counting backward while alerted by a buzzer. By collecting a series of physiological signals and building a model to recognize a pattern, we aim to train a stress and overload detection model to detect unsafe neurological status.

You are invited to participate in this study on the "Assessment of Stress and Overload for Research in Construction." This research aims to detect stress and overload from your physiological signal changes.

Procedures involved in the research: You will be given extreme consideration to protect your health and safety, particularly against COVID-19. You don't need to print this form and COVID LOI as hard copies are available on the experiment day.

You will be given personal protective equipment PPE (i.e. head, hand, hearing, and eye protection, high-visibility hat and vest, and harness lanyard) to wear for the session. You will be asked to bring and wear long pants and winter boots for the day of the experiment. For data collection purposes, you will be given a wristband sensor to wear for the whole duration of the test. A headphone, tablet, pen and paper, and chair will be available on the experiment day. The test will be approximately 60 minutes in length. You have the right to ask for a break or any request to feel comfortable through the sessions.

Test Instruments: Here, you can find the list of Personal Protective Equipment PPE, provided for the experiment:

- Head protection (provided)
- Eye Protection (provided)
- Hearing protection (provided)
- High-visibility hat and vest (provided)
- Hand protection (provided)
- Harness lanyard (provided)

Equipment required by participants

- Long pants (required by participants)
- Foot protection (wearing winter boots required by participants)

Here you can find the list of instruments for data collection:

- E4 wristband for data collection (provided)
- Headphone (provided)
- Pen and paper (provided)
- Chair (provided)
- Tablet for Stroop Colour Word and Race Speed Test (provided)

Activity Guides: The experiment will occur at McMaster's main campus in the open area on the northwest side of University Hall. A few simple questions regarding your gender, age, height, weight, feeling and hour of sleep will be asked with your permission. Then, you are expected to perform the following activities in one session.

1. Sit on a chair and relax for 5 minutes and then perform a ruler test,
2. Stand on the right leg and try to maintain balance for 2 minutes, then relax for 1 minute, then stand on the left leg and try to maintain balance for 2 minutes,
3. Perform the ruler test and relax for 5 minutes
4. Walking on a defined path for 3 minutes
5. Ruler test and Relax for 5 minutes
6. Race speed test performed by tablet for 3 minutes
7. Ruler test and Relax for 5 minutes
8. Stroop colour word test performed by tablet for 3 minutes
9. Ruler test and Relax for 5 minutes
10. Counting backward for 3 minutes
11. Ruler test

Potential harms, risks or discomfort: We make sure that you mindfully participate, and you have the option to leave at any moment. You will be prepared for each task and given five minutes of relaxing time between activities to alleviate the potential pressure and risks. It should be noted that you will be requested to perform actions that aim to trigger mental stress and overload, leading to frustration and requiring patience and focus from your side. These prompts could potentially cause you to be worried or upset. For such activities, you can skip that step or opt out of the study at any time up to three weeks from the data collection day.

Experiments will be conducted on days without snow or rain for your convenience, in addition to removing existing snow or dirt before your arrival. However, there are physical risks associated with this experiment, as they are performed outdoors.

Potential benefits: The research will not benefit you directly. The research goals and objectives that decrease accidents and unsafe human behaviour will help society and the construction industry in many ways.

Incentives: As compensation, you will receive the 25CAD Amazon Gift Card (e-gift card) after participating in the experiment. You will receive a 12.5CAD Gift Card (e-gift card) for each half hour of participation. You will also be reimbursed for parking for the experiment's duration and half an hour.

Confidentiality: All records of the documented observations and individual and recorded database information will be kept private. Information from this experiment only is available to the research supervisor. Descriptions and necessary demographic details will be used only for research purposes. The identification code will be assigned to your recorded information from the moment of data collection instead of your information. The cross-code will be kept in secure storage with

secure access. The given code will be used in all publications, and no one will be identified by name in this study.

Participation and withdrawal: If you feel uncomfortable with the observations, you are welcome to inform the responsible investigator immediately. Every effort will be made to respect your privacy and comfort. Suppose there are any activities or questions you feel uncomfortable going through or would prefer not to contribute. In that case, you may skip over that section or terminate the experiment. Participation in this research is entirely voluntary.

You can withdraw from this study for up to three weeks from the data collection day. In the case of withdrawal, you will receive $12.50 if you withdraw in the first 30 minutes of participation and $25 if you withdraw after 30 minutes of participation. This is due to the incentive of 12.5$ for every 30 minutes of participation.

Information about the Study Results: This study is expected to be completed around the beginning of 2021. If you are interested to receive a summary of the results, please let us know to send it to you.

Questions about the Study: If you have any questions or like to know more about the study, please email the responsible investigator at: eskandah@mcmaster.ca

This study has been reviewed by the McMaster University Research Ethics Board and successfully achieved ethics clearance. Please use the following contact in a case you have concerns or questions about either your rights or about the way the study is conducted:

McMaster Research Ethics Secretariat

Telephone: (905) 525-9140 ext. 23142

C/o Research Office for Administrative Development and Support

Chapter 5 Anomaly Detection Using Unsupervised Deep Learning

5.1 Chapter Overview

This chapter presents a deep learning-based anomaly detection technique developed to identify workers' stress and overload. Due to the nature of the stress and overload, all three levels of acquired data, discussed in Chapter 4 section 4.3, are processed to generate an input to the anomaly detection method. After pre-processed stages are discussed in section 4.4, correlation matrices are produced, presented in section 5.2. The proposed method architecture is given in section 5.3, discussing different areas of the method and its function. After that, in section 5.4, the training process is presented, followed by validation results over two test datasets in section 5.5, and finally, conclusions and limitations in section 5.6

5.2 Input Correlation Matrices

In multivariate data, correlations provide a metric to indicate whether two variables are strongly dependent or not. The Signature Matrix or Correlation Matrix (M^t), presents the correlation between different data pairs and is critical to characterize the system status. In the proposed anomaly detection method, M^t, or pairwise correlations of features, are used as input. Signature Matrices have 14 rows and 14 columns (coming from 14 recorded features).

For a given time step t, the signature matrix (inter-correlation between different pairs of features i, j) is calculated using Eq. (5) to Eq. (7). Window length, referred to as ω, is a period that all readings in that segment are considered for one sample data (such as $[t - \omega, t]$). The following equation is formulated to a $t - \omega$ to t segment for features i and j. κ in Eq. (7) is the rescaling factor representing three different window lengths. By considering $\kappa = \omega$, short-term, medium-term, and long-term data simultaneously can be captured.

$$X_i^{\omega} = (x_i^{t-\omega}, x_i^{t-\omega-1}, \dots, x_i^{t}) \quad \text{Eq. (5)}$$

$$X_j^{\omega} = \left(x_j^{t-\omega}, x_j^{t-\omega-1}, \dots, x_j^{t}\right) \quad \text{Eq. (6)}$$

$$m_{ij}^{t} = \frac{\sum_{\delta=0}^{\omega} x_i^{t-\delta} x_j^{t-\delta}}{\kappa}, \quad m_{ij}^{t} \in M^{t} \quad \text{Eq. (7)}$$

Fig. 14 displays the signature matrix from one subsequent series in time step 50 ($t = 50$), containing data from five previous steps ($\omega = 5$). This matric reflects information recorded from time steps 45 to 50.

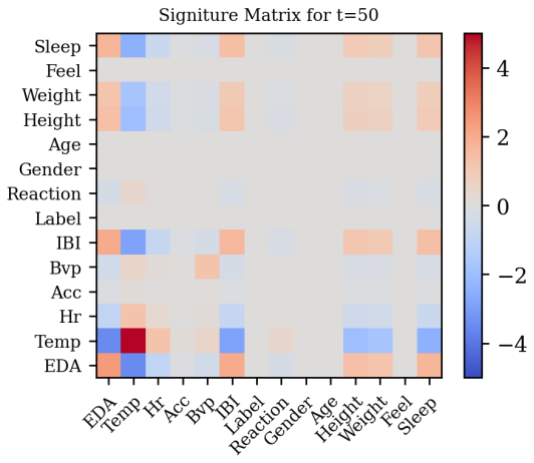

Fig. 14. signature matrix sample for t = 50

To create training and test data for the anomaly detection method, a fixed-length window size ω, slides over recorded data for a length of h steps to generate a set of subsequent multivariate series. Fig. 15 shows a sample of a rolling window on generic data (e.g., x). The rolling window will capture h consecutive readings of signal with a length of ω, those results in a sample data size $h \times x \times \omega$. The same concept applies to signature matrices from recorded 14 features with 14×14 sizes at each timestep. So, the result will be the size of $h \times 14 \times 14 \times \omega$.

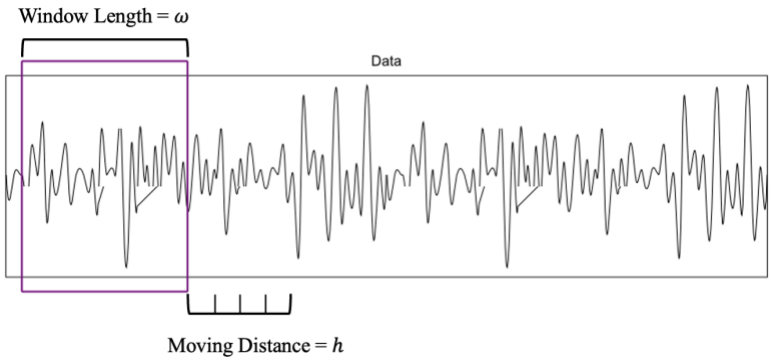

Fig. 15 Sample of a rolling window on generic data

After considering different h sizes in the hyperparameter tuning stage, h equal to 5 has been chosen as an optimum size, which will be presented in section 5.5.4. in. other words, signature matrices of five previous periods or timesteps (each timestep is 0.25 sec considering 4 Hz frequency) are stacked together to reflect the changes over time. Three different sizes of ω are considered for generating data using correlation equation Eq. (5) to Eq. (7), and matrices are stacked in groups of three (i.e., different window sizes of 5, 10, 20). Other window sizes are considered to simultaneously capture short-, medium-, and long-term data, equivalents to RGB channels in colour image application problems. Also, it helps the method to capture different window sizes and have less bias over the selected ω. Ultimately, the input dimensions are (5, 14, 14, 3). A sample

stack of matrices is visualized in Fig. 16. The generated stack of signature matrices is passed along as an input. More details on correlation matrix calculation can be found in Deep Learning Jupyter Notebook (Eskandar, 2022b).

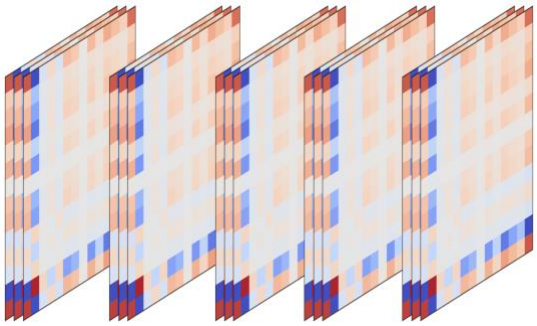

Fig. 16. A sample input with a size equal to (5,14,14,3)

5.3 Anomaly detection method Structure

Section 2.5 compared different anomaly detection techniques applied over multivariant time series. Multi-Scale Convolutional Recurrent Encoder-Decoder (MSCRED), a deep learning autoencoder with LSTM and an Attention layer, is preferred to detect multivariant time series anomalies. Autoencoders are deep learning methods used to reconstruct the same input. Autoencoder labels an input as an anomaly if the reconstruction error is high, in which reconstruction error refers to a difference between the original input and reconstructed output.

Fig. 17 presents the proposed architecture, a deep learning method with several hidden layers and three general sections; Encoder, LSTM, and Decoder explained in sections 5.3.1, 5.3.2, and 5.3.3. Input to the following architecture is a stack of correlation matrices, depicted in Fig. 16. As described in section 5.2, the input includes the correlation between 14 features for five consecutive time steps within three different window lengths of 5, 10, and 20. Consequently, resulting in an

input size of (5,14,14,3). The output of this method is a reconstructed correlation matrix for the last timestep with a flow of information similar to the arrows in Fig. 17.

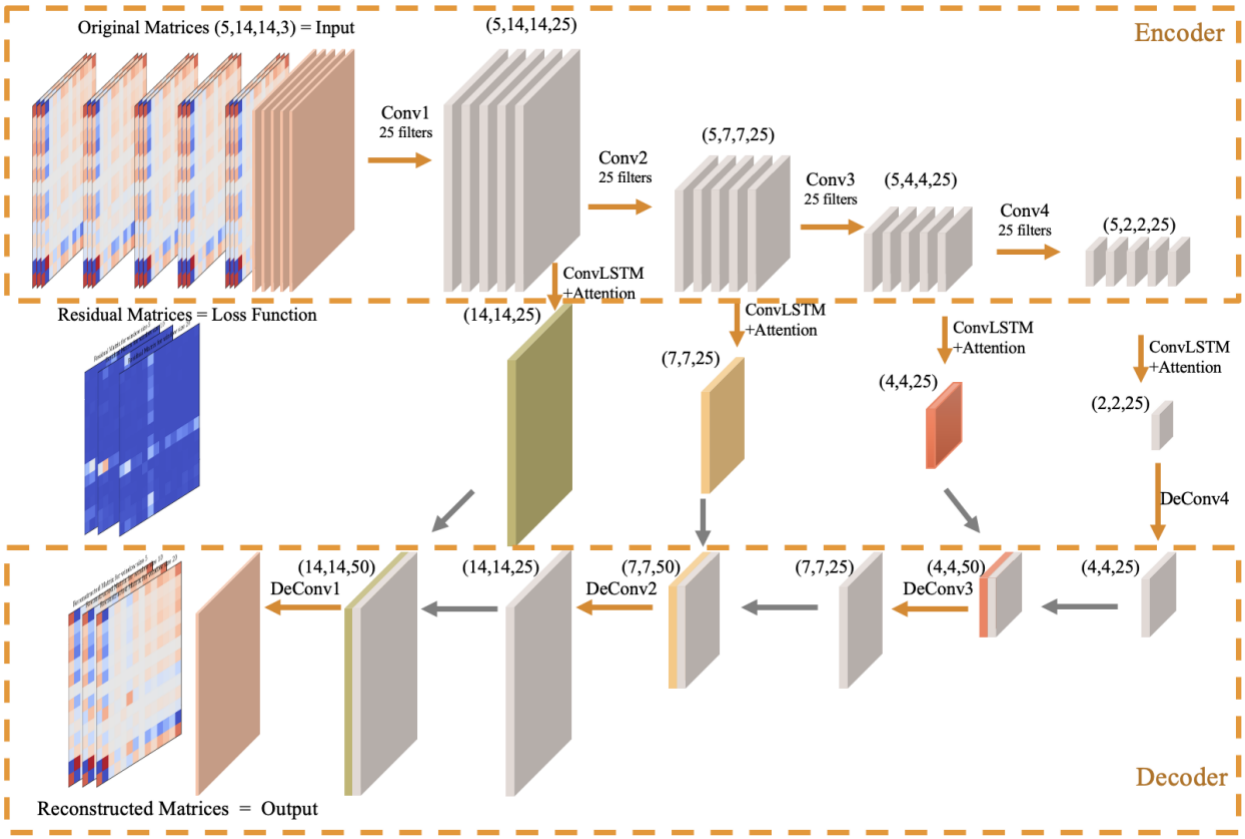

Fig. 17. Anomaly detection method architecture

The network is trained using the TensorFlow library on a Tesla P100-PCIE-16GB GPU cloud processing server. More information on training is presented in section 5.4. All algorithms and libraries used for the anomaly detection method using Python language programming can be found in the Deep Learning Jupyter Notebook (Eskandar, 2022b). Below, different parts of the method and their primary purpose are given.

Fig. 17 demonstrates the detail of the final anomaly detection method architecture. The above optimum method (which leads to lower validation loss) is selected by tuning hyperparameters

through the grid search method. A hyperparameter is a parameter that controls the learning process, such as the number of hidden layers, the size of each hidden layer, or the optimizer algorithm. In this process, different variations for hyperparameters are considered (e.g., three encoding layers instead of 4 or smaller filter sizes or larger strides). Grid search is a simple algorithm for finding the optimum hyperparameter. The search area of the hyperparameters is divided into discrete grids, similar to Fig. 18. All the possible combinations are considered, and performance metrics are calculated to find the optimum solution.

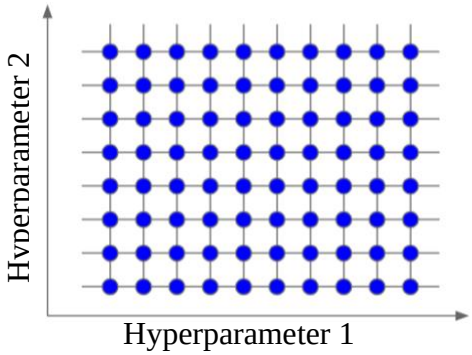

Fig. 18 Grid search algorithm example for hyperparameter tuning

The following sections clarify the method's three main components: CNN Encoder, Convolutional LSTM, and CNN Decoder.

5.3.1 CNN Encoder

The Convolutional Neural Network (CNN) encoder consists of four CNN layers (Conv1 to 4) that reduce the input dimensions by applying filters and extracting features from the input. Conv in refers to a CNN which passes multiple filters across the matrix to generate new features. By applying Conv, the matrix size shrinks based on filter size. 25 filters (kernels of size 2×2 with 1×1 strides) are used in each layer for feature extraction while reducing the size of matrices.

5.3.2 Convolutional-LSTM with Attention

Since the input is time-related, LSTM is added to learn a representation of sequence data, as it can extract the effect of past events (both long-term and short-term effects). An attention mechanism is employed to overcome a shortcoming of the Autoencoder architecture on long sequences. The attention layer is beneficial for capturing spatial features and patterns lost throughout the autoencoder. After each encoder layer, Convolutional-LSTM with attention is added to extract temporal information.

One Convolutional-LSTM is assigned for each CNN layer (tagged as ConvLSTM + Attention) while many-to-one architecture is selected for the middle LSTM layer. The attention layer is an added weight that helps to look at the input sequence's essential parts rather than the whole series.

5.3.3 CNN Decoder

CNN Decoder reconstructs the signature matrix using (DeConv1 to 4 in Fig. 17). CNN Decoder acts in the reverse order of the CNN Encoder by increasing the dimensions of matrices and reconstructing the original matrices. DeConv in Fig. 17 refers to a CNN Decoder that tries to reconstruct the original matrix and up-sample layers. The system architecture is based on an encoder-decoder, and every decoder corresponds to a pooling unit of an encoder.

5.4 Training Process

The dataset for training consists of recorded data from 17 subjects after preprocessing step, while the associated data related to one participant (Participant No. 18) is kept for the test dataset. Training dataset size has the size of (33659, 5, 14, 14, 3), of which 10 percent is kept unseen and

allocated to the validation dataset. Having 10 percent of training data for validation allows the method to monitor the generalizing performance over unseen data after each iteration.

The method contains a total of 116,028 parameters to be trained over a maximum of 150 epochs/iterations and improve itself and lower the loss. In the training process, achieving high accuracy on the training dataset does not lead to developing a method that generalizes well over the test dataset. After each iteration, the training loss and validation loss are calculated based on the formula presented below, Eq. (8). The method should terminate the training process if it cannot improve the validation loss after a certain number of consecutive iterations. This technique is called early stopping, and it's applied to prevent the method from overfitting the training dataset and avoiding unnecessary training times. Seven iterations are selected for early stopping after several tries and errors. The number of iterations depends heavily on the neural network's size, type, and optimizer. Frequently 3 to 10 consecutive iterations are adopted for the early stopping of similar networks (Hasty.ai, 2022; Terry et al., 2021).

TensorFlow (Abadi et al., 2015) is employed for implementing the model and tuning the parameter. TensorFlow is an open-source software library created for Machine Learning (ML) and artificial intelligence. TensorFlow has a particular focus on the training of deep neural networks and their inference process. The proposed method is trained using a Colab GPU cloud processing server (Colab Pro - Tesla P100-PCIE-16GB) to increase the training speed by roughly 65 times (Yuqi, 2021) instead of using Colab CPU cloud processing (Colab Pro - Intel(R) Xeon(R) CPU @ 2.30GHz). More information on the training processing time can be found in Chapter 7.

Training a method that can lead to a good generalization over unseen data is challenging and requires optimizing parameters based on the validation set loss. To minimize the loss, the Adam

optimizer (Kingma & Ba, 2014) is employed with the mini-batch stochastic gradient descent method. Adam is an optimization algorithm, an extension to a stochastic gradient descent method, based on adaptive estimation of the first-order and second-order moments. In each iteration, step loss is calculated using Eq. (8) as the absolute difference between the original and reconstructed matrix for all items in the training dataset. Fig. 19 presents the method training and validation loss progress for 100 iterations. Until the 100th iteration, the accuracy was improving, and then for seven consecutive iterations, there wasn't any lower value for validation loss.

$$\text{Loss} = \sum_{t=1}^{5}\sum_{k=1}^{3}\left|\left|\chi_{:,:,k}^{t} - \hat{\chi}_{:,:,k}^{t}\right|\right|_{n}^{2} \qquad \text{Eq. (8)}$$

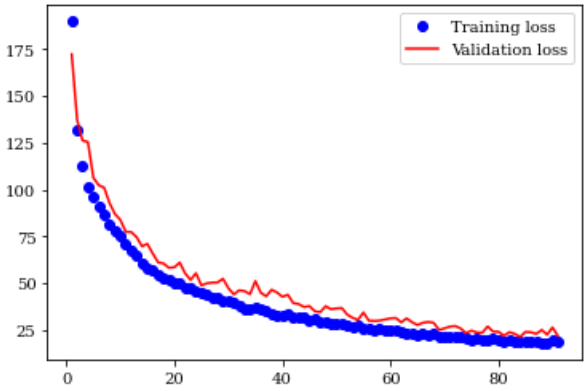

Fig. 19. Training and validation loss

The anomaly detection method output, residual matrix (i.e., reconstruction error), is calculated in the last step by subtracting reconstructed pairwise correlation values from the input pairwise correlation values. If reconstruction error (i.e., the absolute sum of residual values for every feature, step, and window size) is above a calculated threshold, that moment is flagged as an anomaly. The threshold is calculated for each activity taking situation-level data into account. Activity type is considered in the threshold calculation (Relax, one-foot balance, walking, speed

test, Stroop test, and counting backward). A threshold equal to 90 percentiles of the reconstructed error for each activity group is calculated and applied to detect anomalies.

Similar to considering the activity type, other features can be regarded as and form a decision fusion. Additional features for decision fusion can be the application-level data (i.e., type of job, nature of work, and industry). It helps the application calibrate the stress detection results for any specific workplace or occupation. While this phase of the work will remain for the authors' future work, the decision fusion presented in this paper focuses only on calibrating the threshold mentioned above for each activity, considering the application-level data. The level of confidence is set to 90 percentiles of reconstructed error over the test set. In other words, if the reconstructed error is high, the input to the method is ladled as an anomaly, demonstrated in detail later.

5.5 Validation Results and Discussion

This section presents anomaly detection analysis and results by applying the trained method to two test sets. The residual matrix is discussed in the next section as an interpretation tool to comprehend contributing features for detected anomalies.

5.5.1 Residual Matrices

Residual matrices bring insight into the interpretation of contributing features. Fig. 20 illustrates the residual matrix of a standard sample with a reconstruction error less than the selected threshold versus an anomalous sample with a higher reconstruction error than the threshold. The contributing features in detecting an anomaly from a residual matrix are presented with a heat map. The

following figure shows the predicted correlation values for accelerometer (Acc) and heart rate (Hr) are the main contributing features to this anomalous sample.

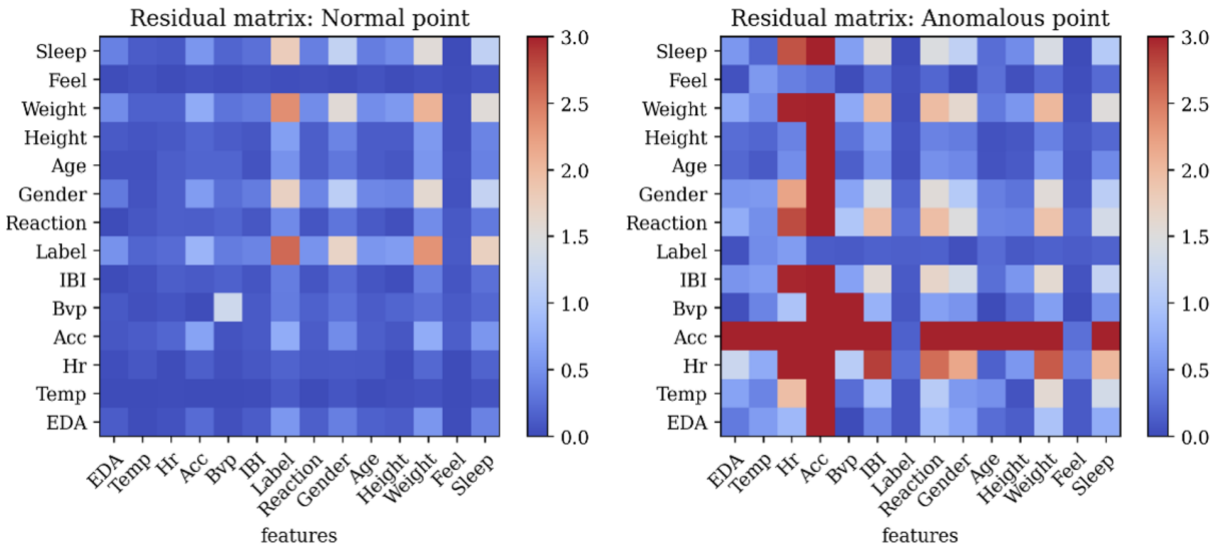

Fig. 20. Residual matrix of a normal sample vs. an anomalous sample

An anomaly can be detected by assessing the number of poorly reconstructed pairwise correlations (i.e., signature matrix) equal to a more significant reconstruction error or greater values in residual signature matrices. A threshold equal to 90 percentiles of the reconstructed error is used to detect anomalies. Precision Eq. (9), Recall Eq. (10), and F1-Score Eq. (11) are used to validate the performance of the anomaly detection method. In the following equations, TP, FP, TN, and FN are True Positive, False Positive, True Negative, and False Negative.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad \text{Eq. (9)}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad \text{Eq. (10)}$$

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision+Recall}} = \frac{\text{TP}}{\text{TP}+\frac{1}{2}(\text{FP}+\text{FN})} \qquad \text{Eq. (11)}$$

Precision Eq. (9) is the proportion of the detected classes that were correctly detected. Precision answers the question of how many detected anomalies were truly anomalous. Recall Eq. (10) is the ability to find all relevant categories, which is the proportion of correctly detected anomalous to all actual anomalies. Moreover, F1-Score Eq. (11) considers both precision and recall and offers a useful metric that is a harmonic average.

5.5.2 Test Dataset I

The first test dataset size is (8211, 5, 14, 14, 3), which includes information captured from one participant out of 18 who participated in the experiment. It is worth repeating that test datasets were kept unseen through training for accurate testing purposes. Anomalies of the test dataset are identified and labelled based on the expert's opinion by looking at the raw collected time series. In other words, valid abnormalities of this dataset were visually recognized by spotting an uncoordinated movement in multivariant time series (e.g., a sudden spike or drop in heart rate while the acceleration is steady). The results of accuracy metrics over the test dataset I are presented in Table 3.

The macro average considers all classes equally, irrespective of the number of support values (i.e., class weight). Macro averaged F1-score is computed using arithmetic mean or unweighted mean. On the other hand, the weighted-averaged F1-score is calculated by calculating the F1-score mean considering each class's support number (i.e., class weight). Accuracy, or micro F1, or global F1-score average, is presented in Eq. (12). To compute accuracy, it first requires calculating the sum of TP, FP, and FN values for all classes, then values are put into the F1-score equation Eq. (11).

$$\text{Accuracy} = \frac{\sum \text{TP}}{\sum \text{TP} + \frac{1}{2}(\sum \text{FP} + \sum \text{FN})} \qquad \text{Eq. (12)}$$

Table 3 presents precision, recall, and F1-score for both classes with support numbers and the number of actual samples in the test dataset. Accuracy, macro averaged, and weighted-averaged F1-score are given, showing that the trained DL-based anomaly detection method provides 98% accuracy.

Table 3. Unsupervised DL-based model accuracy using test dataset I

	Precision	Recall	F1-score	Support
0 Non-Anomaly	1.00	0.98	0.99	8129
1 Anomaly	0.35	1.00	0.52	82
accuracy			0.98	8211
macro avg	0.68	0.99	0.76	8211
weighted avg	0.99	0.98	0.99	8211

The Unsupervised DL-Based method performance over the test dataset I shows 1.00 precision for non-anomalies and 0.35 for anomalies. The method can recall 98% and 100% of cases for non-anomalies and anomalies. F1-Score for anomalies is 0.52, and 0.99 for non-anomalies. Accuracy as a global F1-score average is 0.98, the macro average is 0.76, and the weighted average of 0.99. Based on the defined metrics, the method can predict all anomalies in the first test dataset. However, it is detecting more anomalies than exist due to low precision. It is beneficial in highly sensitive circumstances where the false alarm is accepted for a higher level of safety.

Fig. 21 demonstrates the reconstructed error at each timestep from predicting the test dataset I. Horizontal grey ribbons represent a new activity in the series of actions that the participants of the experiment undertook. Fig. 21, the reconstructed error is presented in a series of connected blue dots with horizontal black lines as the threshold. All data points above the threshold are detected

anomalies, presented with orange stars. Anomalies represent a higher error than the calculated threshold (i.e., 90 percentiles of error for each activity).

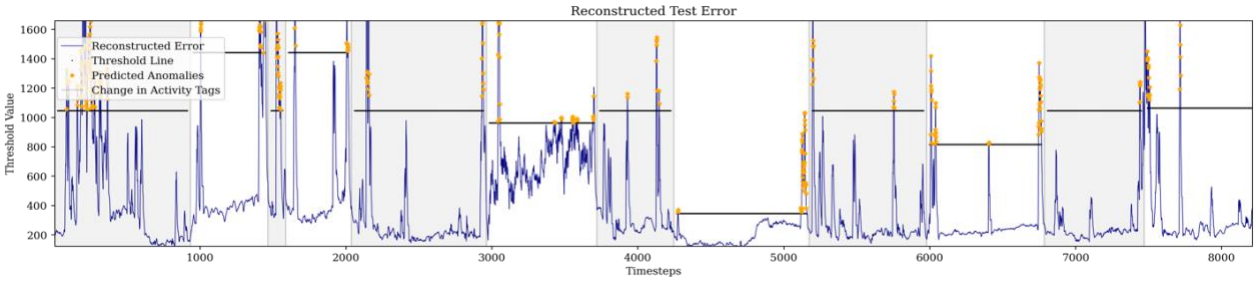

Fig. 21. Reconstruction error on the test set and detected anomalies

A portion of the test dataset timeline, starting from timestep 7470 and ending in 7730, is presented in Fig. 22. This moment is the start of a new activity in the test experiment, in which the participant was asked to count backward by 7. As shown in Fig. 22, the threshold line is applied to start from the timestep 7490 (i.e., 5sec later). The same technique is used for the rest of the datasets when activity changes.

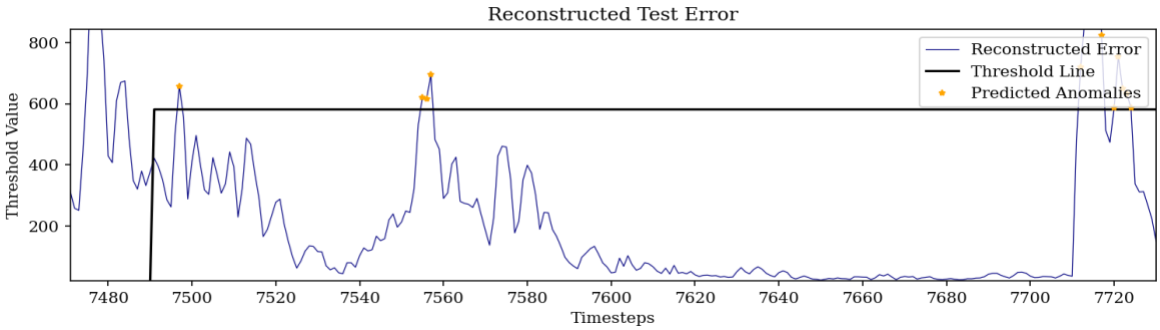

Fig. 22. Visualising anomaly detection results for the last activity in the test dataset I

To reduce false positives, a few seconds after the beginning of each activity, the threshold has changed to a higher number. The threshold value in this zone is set to a higher value to prevent the

method from detecting false positive anomalies. It is probable to detect anomalies when participants were asked to change the activity and instruction was given for what to do due to rapid change of the existing state. It's unknown whether a participant started the next activity in 1 sec or 5 sec, had delayed understanding the instruction and was in a transition state. So, the transition period is removed from the calculation by increasing the threshold to a maximum of five seconds before and after each tag. More details are presented for the mentioned zone in Fig. 23 and 24.

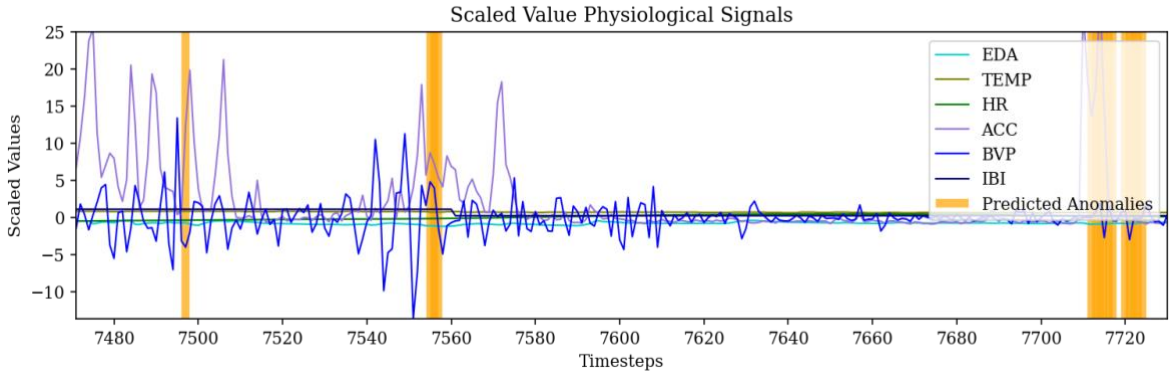

Fig. 23. Feature values for the last activity in the test dataset I

Fig. 24 presents the residual matrix of a standard sample with a reconstruction error less than the selected threshold versus an anomalous sample with a higher reconstruction error than the threshold. We can identify the contributing features in detecting an anomaly by investigating the residual matrix at each timestep. An anomalous sample has a higher reconstruction error than a standard sample, for which the trained method failed to simulate the signature matrix identical to the input. For instance, for time step 7555 in a dataset, by analyzing the residual matrix presented in Fig. 24; we can identify that Acc and BVP signals are the most contributing features in the detected anomaly. The residual matrix at 7535 is a standard sample with insignificant values.

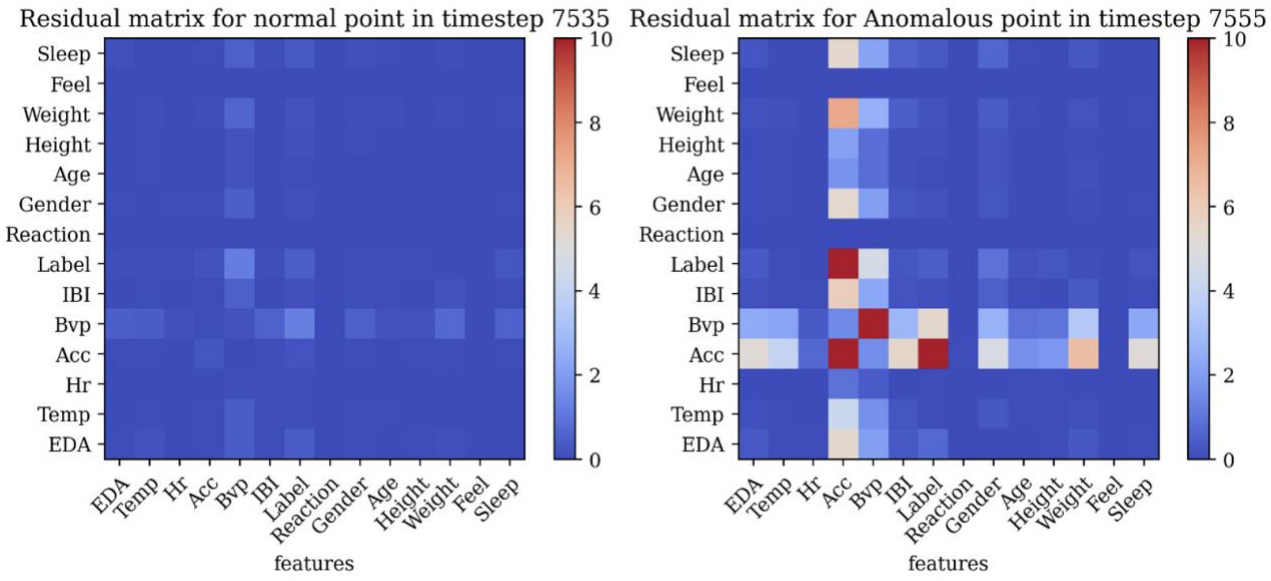

Fig. 24. Residual matrix of a normal (timestep 7535) vs. an anomalous sample (timestep 7555)

5.5.3 Test Dataset II

Another data collection was arranged for testing purposes of the proposed method to label data while collecting. This experiment was conducted without following the original sequence of activities in order to prove the method's independence from the sequence of activities in the designed experiment. While recording data, the moments of intentionally added extra mental stress/overload were logged as labels. Added mental stress and overload are arranged by distracting the participant using an unexpectedly moving object in front of the participants, a buzzer, and sudden talking during the experiment.

The second test dataset size (2895, 5, 14, 14, 3) includes tagged timesteps. While participants got periodically stressed/disturbed/distracted, timesteps are captured in a separate file. Tags on the occurrence of these methods are assumed as valid anomalies that can validate the proposed

unsupervised deep learning method. The results performance of the trained method for predicting anomalies over test-set II is presented in Table 4.

Table 4. Unsupervised DL-based method accuracy using test dataset II

	Precision	Recall	F1-score	Support
0 Non-Anomaly	0.97	0.92	0.94	2712
1 Anomaly	0.33	0.57	0.42	183
accuracy			0.90	2895
macro avg	0.65	0.75	0.68	2895
weighted avg	0.93	0.90	0.91	2895

The Unsupervised DL-Based method performance over the test dataset II shows 0.97 precision for non-anomalies and 0.33 for anomalies. The method can recall 92% and 57% of cases for non-anomalies and anomalies. F1-Score for anomalies is 0.42, and 0.94 for non-anomalies. Accuracy as a global F1-score average is 0.90, the macro average is 0.68, and the weighted average of 0.91. The given performance metrics don't discuss the temporal aspect of the problem. So, by examining samples over the timesteps and analyzing each detected anomaly, the method performance can be observed in a better way. Reconstructed errors and indicated anomalies over time are illustrated in Fig. 25 and Fig. 26.

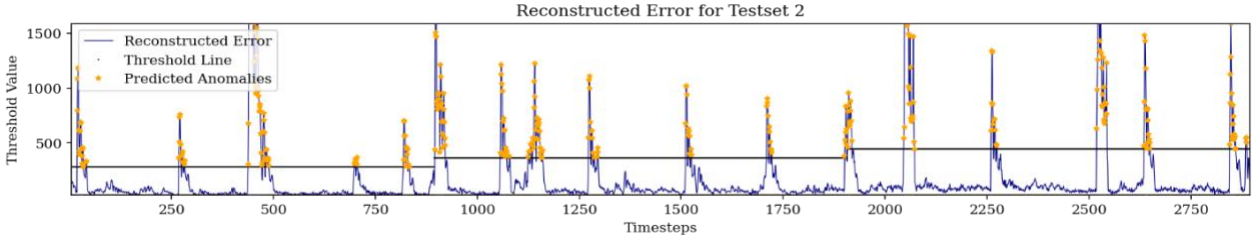

Fig. 25. Reconstruction error and detected anomalies on the test dataset II

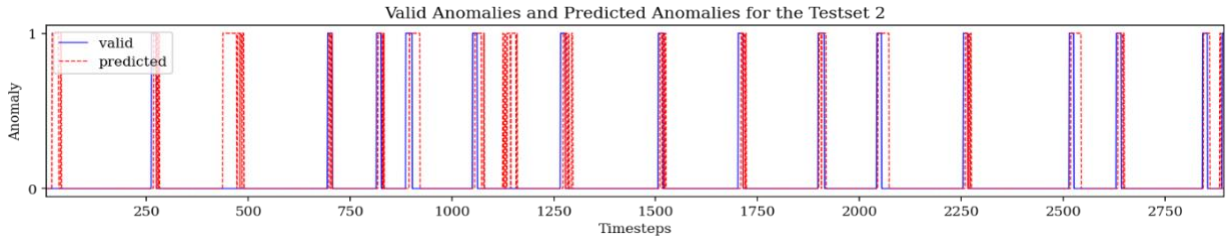

Fig. 26. Valid anomalies and predicted anomalies for the test dataset II

As Fig. 26 presents, the moments of valid and predicted anomalies match in most cases, prediction starts slightly later, and it predicts more anomalies a few seconds later. Two moments during the experiment, the method detected the anomalies that weren't tagged during the experiment. For a portion of the above test set (timestep 1200 to 1900), details are presented in Fig. 27 and Fig. 28.

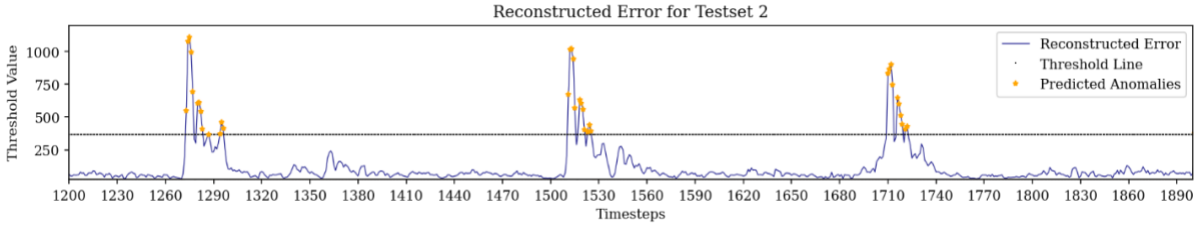

Fig. 27. Part of reconstruction error and detected anomalies on the test dataset II

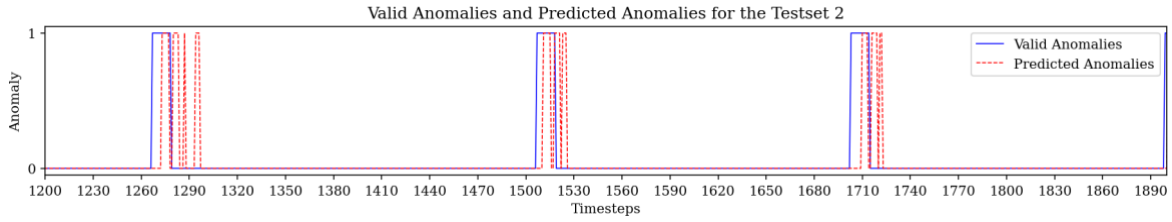

Fig. 28. Part of valid anomalies and predicted anomalies for the test dataset II

As Fig. 28 presents, there are moments of the predicted anomalies after each valid anomaly that are considered false positive, however, the tagged labels are not precise. From this portion of test

dataset II, two examples of residual matrices, one standard sample and one anomalous sample, are presented in Fig. 29.

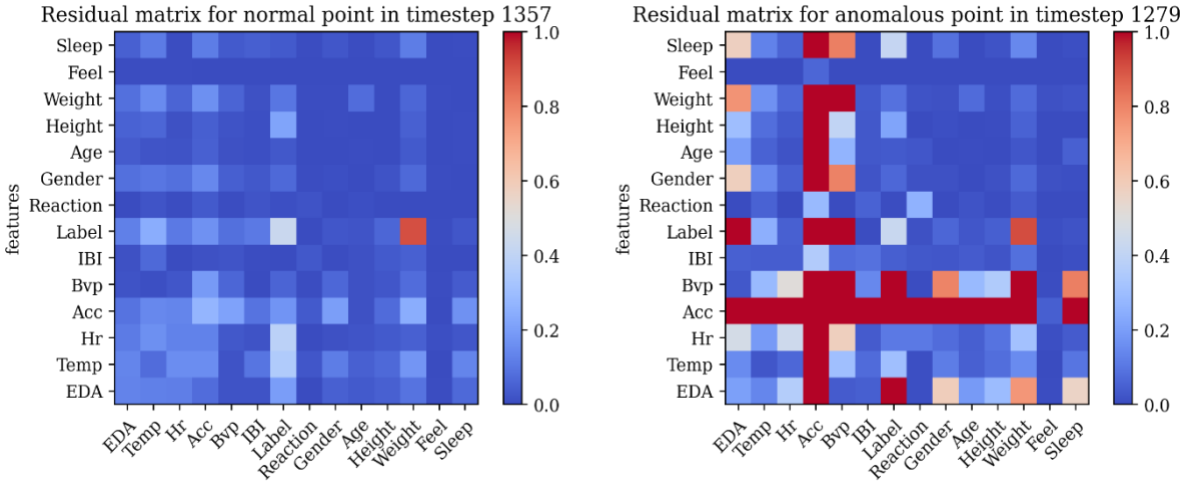

Fig. 29. Residual matrix of a normal sample vs. an anomalous sample

The residual matrix for the anomalous at timesteps 1279 is presented in Fig. 29, which shows an unconventional correlation between accelerometer, blood volume pressure, electrodermal activity, and label. Residual matrices are significant indicators for identifying uncoordinated features that lead to an anomaly.

5.5.4 Window Size Impact

To capture the impact of window size on the performance of the anomaly detection mode, different window sizes are selected for training and testing while keeping the same number of samples. The larger window sizes include a more extended period for the method to capture the trend, resulting in higher calculation time. As demonstrated in

Fig. 30, the optimum window size is five, leading to the least processing time while keeping a low training and validation loss. The processing time is calculated for training 116,028 parameters with training dataset size (33659, 5, 14, 14, 3), using Tesla P100-PCIE-16GB.

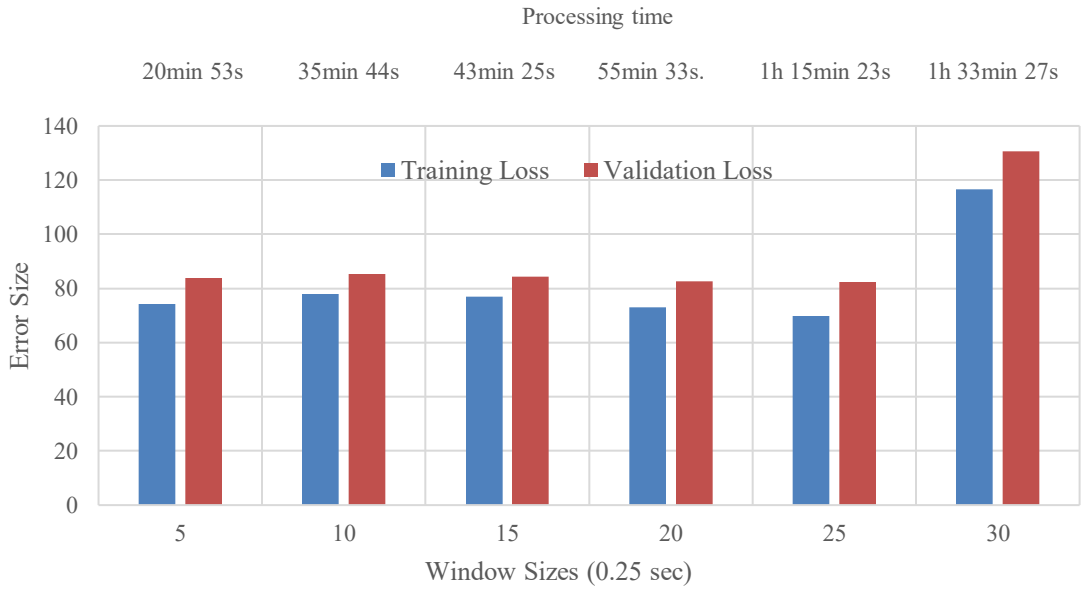

Fig. 30. Anomaly detection method training loss and validation loss for different window sizes

5.6 Conclusion and Limitation

This chapter presented a deep learning-based method for detecting workers' stress and overload based on the acquired data, which was discussed in Chapter 4. All three levels of acquired data are pre-processed, aligning frequencies and time intervals to increase method efficiency, which leads to improved information accuracy and method quality. The proposed method includes unsupervised learning formulated over the correlation of 14 features stacked together for five consecutive timesteps and three different window sizes. Fig. 31. presents the overview of occupational stress detection and contributing feature assessment.

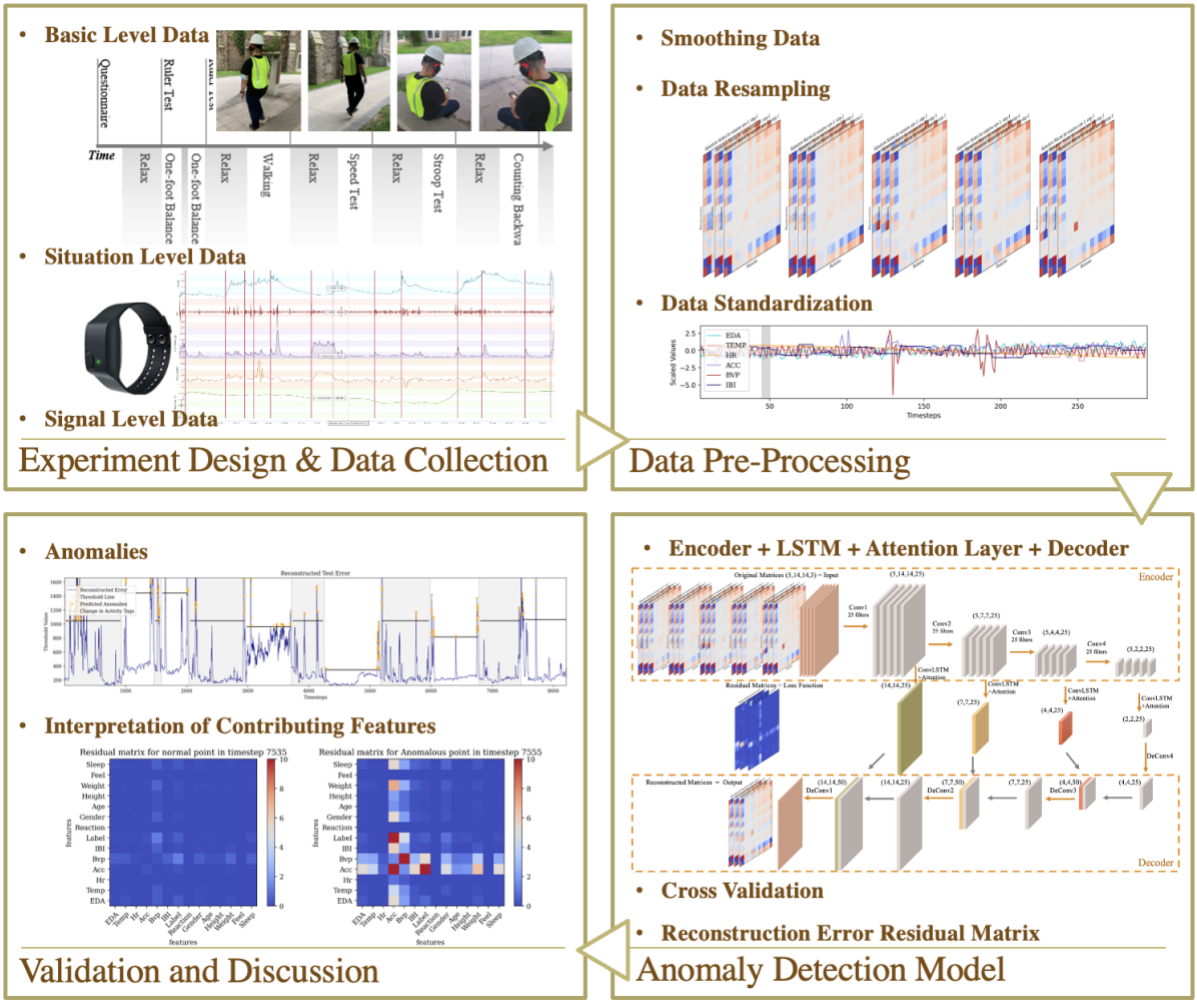

Fig. 31. Deep Learning-Based Anomaly Detection Research Methodology

The framework presents a deep learning-based anomaly detection method for the stress detection model. It contains four sections: experiment design and data collection, data pre-processing, anomaly detection method, and validation and discussion, as explained in the methodology. A wide range of data is collected, including physiological signals using a wearable biosensor. The collected data are labelled and sliced into smaller sections stacked together to record the changes over time. Then, an unsupervised deep learning anomaly detection was developed that identifies

patterns and reconstructs the input. High reconstruction error indicates the method's inability to predict the pairwise correlation (by studying at the previous 1.25 seconds), flagging an anomaly. The method was tested against two separate test datasets resulting in 98 percent accuracy against the first test dataset. This test dataset was tagged by the expert's opinion by looking at the raw collected time series. The second test dataset was arranged without following the original sequence of activities in order to prove the method's independence from the sequence of activities. While recording data, the moments of added extra mental stress/overload were logged as labels. So, the second test dataset includes tagged timesteps based on real-time observations of incidents. The method showed 90 percent accuracy against the second test dataset.

Despite the high accuracy of the developed method, assuming that the outcomes can be compared against absolute truth in a problem space uncertain by nature is unrealistic. Therefore, the stress detection method discussed in this chapter suffers from the lack of availability of absolute truth. This method identifies anomalies when the reconstruction error exceeds the defined threshold. Defining a higher threshold will result in high false negatives, while a lower threshold value leads to high false positives. Moreover, training unsupervised deep learning methods, such as discussed in this chapter, requires thousands of training samples due to a higher number of parameters to be calculated. The training time for DL-based anomaly detection is high, which leads to a higher cost of training in addition to a higher cost of data collection. So, setting thresholds, over-reliance on absolute truth, and the need for countless training samples is this method's shortcoming. To overcome the above-mentioned challenges and shortcomings, further research and investigation resulted in the designing and development of an adaptive Neuro-Fuzzy Inference System for anomaly detection.

Chapter 6 Anomaly Detection Using Adaptive Neuro-Fuzzy Inference System

6.1 Chapter Overview

The previous chapter proposed a method to detect anomalous states of stress using deep neural networks. Looking into the limitations of the proposed method, such as setting a threshold for detecting anomalies and the need for thousands of training samples, motivated this research to investigate a practice using fuzzy logic. Although Artificial Neural Networks (ANN) enable self-learning, the absence of human inference makes such a method challenging to comprehend. Besides, it consumes thousands of samples to train, which makes it inefficient. In contrast, the Fuzzy Inference System (FIS) is fast and easy to interpret compared to neural networks. In particular, the Fuzzy set theory attempts to consider human reasoning by using approximate and uncertain information in making decisions. In cases of uncertainties in decision-making, Fuzzy logic improves the system's performance by assigning a feature value between zero and one rather than binary values.

The fuzzy inference system offers promises to address the challenges and shortcomings of the Artificial Neural Network stress detection method, which was proposed in Chapter 5. The problem space in this research, which is to identify abnormal psychological states using physiological info, is inherently vague and fuzzy. Unlike high-level ANN, fuzzy logic algorithms are based on lower-level machine control methodologies.

This chapter presents a proof of concept for the potential use of FIS methods to detect the degree of anomaly. Adaptive Neuro-Fuzzy Inference Systems (ANFIS), a form of FIS-based method, is selected, in which the self-learning ability of ANN and the linguistic expression function of FIS are combined. In this method, the input, output, and rules are modelled as neurons, in which the values of fuzzy sets are transferred as weights.

The organization of this chapter is as follows. The next two sections describe the Adaptive Neuro-Fuzzy Inference Systems (ANFIS) and their architecture. Section 6.4 describes the detail of the modified ANFIS for the anomaly detection problem in this thesis. Then, section 6.5 describes experimental results and discussion, and the last section presents the conclusion.

6.2 Background

Fuzzy logic promotes solving problems with an imprecise spectrum of data that enables finding an array of accurate conclusions. A fuzzy Inference System (FIS) typically contains four stages: (1) fuzzification, (2) rules base, (3) inference, and (4) defuzzification.

1. Fuzzification refers to the process of applying membership functions (MF) to convert crisp input values into a degree of membership of fuzzy sets. An MF for a fuzzy set is defined as $\mu(x) \rightarrow [0,1]$, where each element of x is mapped to a value in the range of $[0,1]$, called membership value or degree of membership.
2. Fuzzy rule base stores IF-Then rules, which are the natural language-based logics defined by experts.
3. The inference results from applying fuzzy set input in the fuzzy rules base.
4. Defuzzification converts the fuzzy set conclusion into crisp output.

There are two major kinds of FIS, Mamdani FIS (Mamdani & Assilian, 1975) and Sugeno FIS (Takagi & Sugeno, 1985). These two are different in the last stage (i.e., defuzzification). The former uses techniques such as the center of gravity to obtain the crisp output, and the latter uses a weighted average to calculate the crisp outputs.

Traditional fuzzy systems mainly depend on the understanding and knowledge of the problem to acquire an acceptable prediction result, especially in case of incomplete or insufficient information. An adaptive neuro-fuzzy inference system (ANFIS) was developed by Jyh-Shing Roger Jang (1997) to solve fuzzy inference system limitations. ANFIS is a form of FIS based on the Sugeno model (Takagi & Sugeno, 1985), which combines the self-learning ability of ANN and the linguistic expression function of fuzzy inference. In this method, the input, output, and rules are modelled as neurons, in which the values of fuzzy sets are transferred as weights. Since membership functions determined by humans are rarely optimum, fine-tuning the membership functions is recommended when holding a large data set. In ANFIS, membership functions and fuzzy rules are calculated from existing data instead of expert knowledge or intuition. This adaptive method provides the system with the benefit of both ANN and fuzzy, making it a classification method. The rule base of a neuro-fuzzy inference system is constructed as a neural network.

Fuzzy control is the most widely used application of fuzzy set theory and fuzzy inference systems. The adaptive capability of ANFIS makes it almost directly applicable to adaptive control and learning control. The nonlinearity and structured knowledge representation of ANFIS are its primary advantages over classical linear approaches. The wide use of ANFIS in signal processing

supports this argument (Depari et al., 2007; Dogantekin et al., 2010; Kamel et al., 2009; Lei et al., 2007).

6.3 ANFIS Architecture

ANFIS architecture, explained by Jyh-Shing Roger Jang (1997), is depicted in Fig. 32 which contains five layers, each responsible for one stage of adapting a fuzzy system. The functions of these layers are fuzzification, applying rules, a sum of rules/normalization, adaptive nodes, and calculating the overall output.

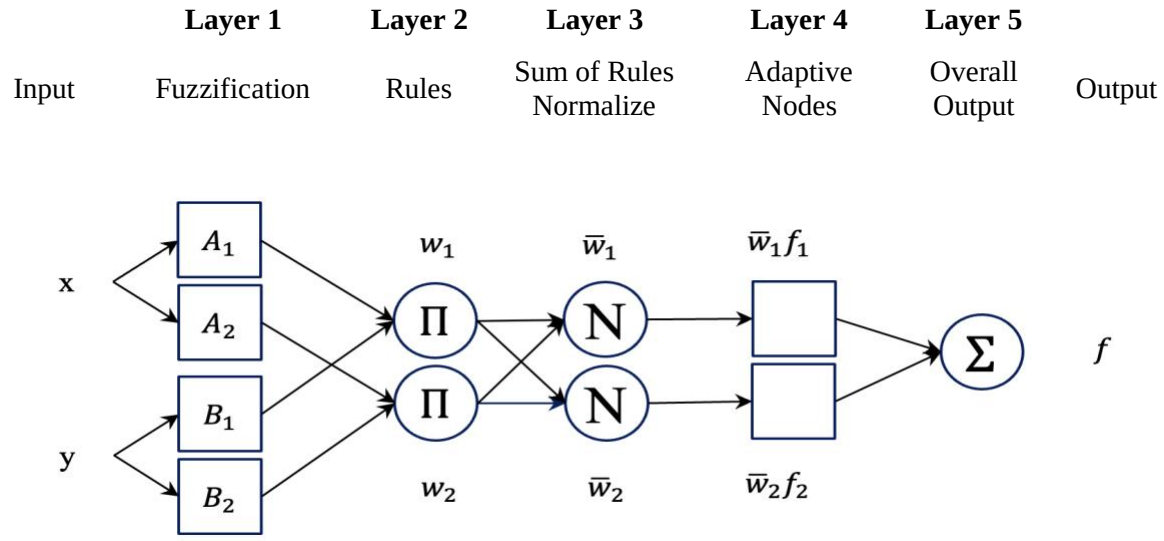

Fig. 32: ANFIS architecture for two inputs

Layer 1: The input (crisp data) passes through the first layer employing the membership function. Nodes in this layer are adaptive nodes and incorporate a membership function quantifier of a fuzzy set. Adaptive nodes in neural networks are improving the model's performance (e.g., prediction and pattern recognition) by altering themselves (change of membership function parameter here) to the latest optimal structure and changing inputs to the next layer while training.

Each node in this layer indicates the degree to which the input x (or y) satisfies the quantifier A (or B). To some extent, A_i is a linguistic label related to a node (e.g., i). For instance, A_1 can be an adaptive node for a $\mu_{A_1}(x)$ membership function that indicates the degree to which x is small, and A_2 incorporates $\mu_{A_2}(x)$ membership function that indicates the degree to which x is large. The quantifier, $\mu(x)$, can be any MF, such as the generalized bell shape Eq. (13).

$$\mu_A(x) = \frac{1}{1+\left|\frac{x-c_i}{a_i}\right|^{2b}}. \qquad \text{Eq. (13)}$$

Parameter set in MF, $\{a_i,\ b_i,\ c_i\}$, change and include broad forms of MF for a fuzzy set. These parameters in layer one are called premise parameters that enable the adaptive function of the first layer. A range of possible values is assigned to each parameter. Applying an optimization algorithm such as Particle Swarm Optimization, the optimum value for each parameter is calculated from the range, which is explained in detail in section 6.4.5.

Layer 2: This layer contains fixed nodes that export the product of all incoming values following Eq. (14).The number of equations in this layer is equal to the product of the number of nodes in the previous layer.

$$W_i = \mu_{A_i}(x)\mu_{B_i}(y) \qquad \text{Eq. (14)}$$

These nodes perform the fuzzy T-norm (triangular norm) operator, aggregating two membership grades as Eq. (15). More information on different operators can be found in the “Neuro-Fuzzy and Soft Computing” by Jyh-Shing Roger Jang (1997).

$$\mu_{A\cap B}(x) = T(\mu_A(x), \mu_B(x)) = \mu_A(x) * \mu_B(x) \qquad \text{Eq. (15)}$$

Layer 3: Each node in layer 3 is a fixed node that calculates the ratio of rules Eq. (16):

$$\bar{w}_i = \frac{w_i}{\sum_i w_i} \quad \text{Eq. (16)}$$

Layer 4: The rule system in ANFIS follows an adaptive function. A common rule set with two fuzzy if-then rules follows Eq. (17):

$$\text{Rule1: if x is } A_1 \text{ and y is } B_1 \text{ then } f_1 = p_1x + q_1y + r_1 \quad \text{Eq. (17)}$$

$$\text{Rule2: if x is } A_2 \text{ and y is } B_2 \text{ then } f_2 = p_2x + q_2y + r_2$$

Like layer 2, nodes in this layer are adaptive nodes following Eq. (18) with a $\{p_i, q_i, r_i\}$ parameter set, referred to as consequent parameters.

$$\bar{w}_i f_i = \bar{w}_i(p_i x + q_i y + r_i) \quad \text{Eq. (18)}$$

Layer 5: This node is a fixed node responsible for calculating the overall output, which calculates the summation of all incoming values Eq. (19).

$$f = \sum_i \bar{w}_i f_i = \bar{w}_1 f_1 + \bar{w}_2 f_2 \quad \text{Eq. (19)}$$

Fig. 33 contains a detailed view of the adaptive neural fuzzy inference system with two inputs.

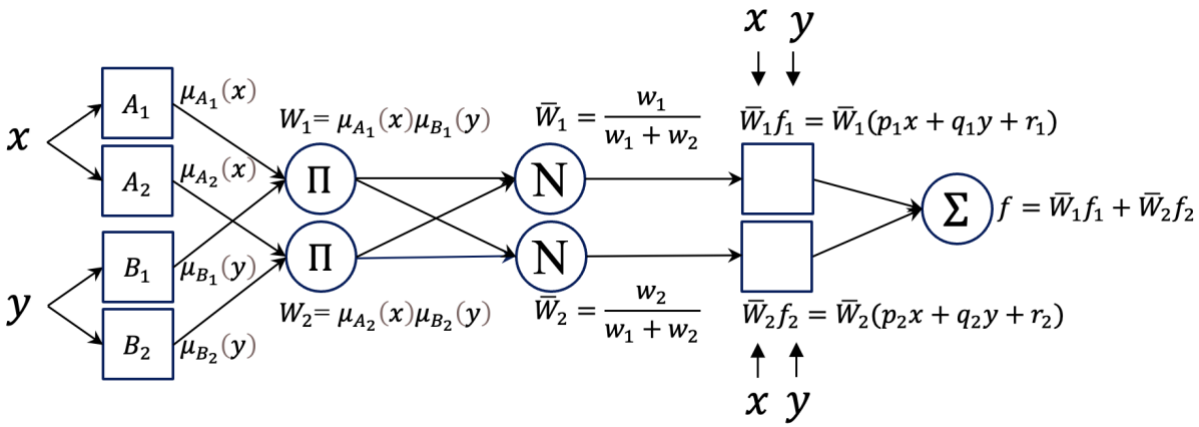

Fig. 33 Detailed ANFIS architecture

With the architecture, as mentioned earlier, two sets of parameters exist. A hybrid learning algorithm can be applied to calculate the premise parameters (nonlinear) and consequent parameters (linear). During the forward pass, consequent parameters in layer four are computed using the least-squares method, and the network error is propagated backward to update the premise parameters in layer 2 using the gradient descent method.

The basics of the ANFIS system, as presented above, can be modified and replaced by other parameters. For instance, the generalized bell shape membership function, which was explained in layer 2, can be replaced by any other type of membership function such as (A) triangular, (B) z-shape, (C) trapezoidal, (D) s-shape, (E) sigmoid, and (F) Gaussian (Fig. 34)

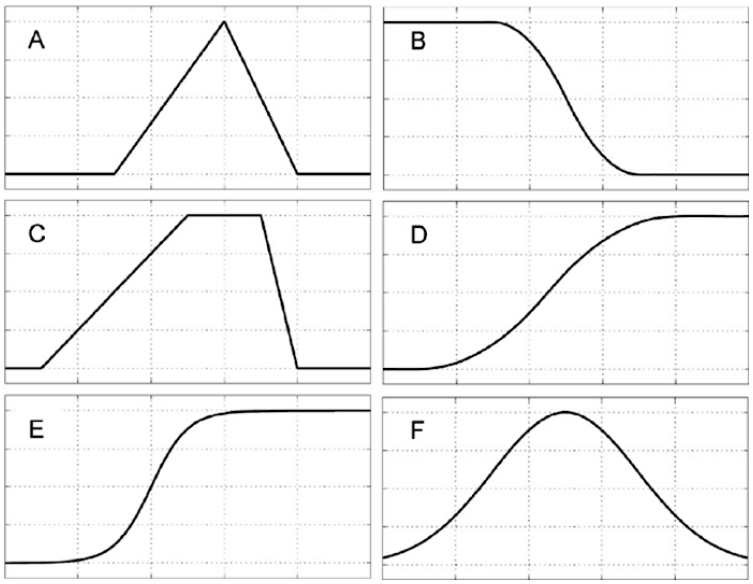

Fig. 34: Membership function (A) triangular, (B) z-shape, (C) trapezoidal, (D) s-shape, (E) sigmoid, (F) Gaussian

Likewise, different optimization algorithms can be applied to calculate the nonlinear parameter in layer two and linear parameters in layer 4. As Palupi et al. (2016) proposed, combining particle

swarm optimization with ANFIS leads to a better result in terms of interpretability and accuracy for classification problems. Similarly, the ANFIS-PSO combination has shown promising results (Alarifi et al., 2016; Robati & Iranmanesh, 2020).

Finding the proper structure is critical to the successful application, including finding an appropriate input-space partition style, which is the number of membership functions for each input. For instance, x and y in Fig. 33, each hold two partitions: A_1 , A_2 for x and B_1 , B_2 for y.

The importance of detecting proper structure increases as the input dataset gets larger. Choosing an effective partitioning system for the input space reduces the number of rules and makes the process faster in both the learning and application phases.

6.4 Proposed ANFIS for Stress Detection

This section represents details on ANFIS implementation. Fig. 35 illustrates the general architecture of the method. The rule base of a neuro-fuzzy inference system is constructed as a neural network. The input, output, and rules are modelled as neurons, in which the values of fuzzy sets are transferred as weights. Each layer of network performs a similar function to section 6.3.

6.4.1 Input Data

Input to the method is the six physiological signals and labels of the performed activities. A window size of 5 readings with a frequency of 4 Hz (i.e., 1.25 sec), similar to the first method, is considered. Having a limitation of one input in the fuzzy inference system for each feature prevents us from choosing time series as an input to this method. To address this limitation, the proposed

method uses changes in values over a 1.25-sec window size as input. The changes (Delta) are calculated using the absolute value of the minimum and maximum (Eq. (20)).

$$\Delta = |\max(x) - \min(x)| \qquad \text{Eq. (20)}$$

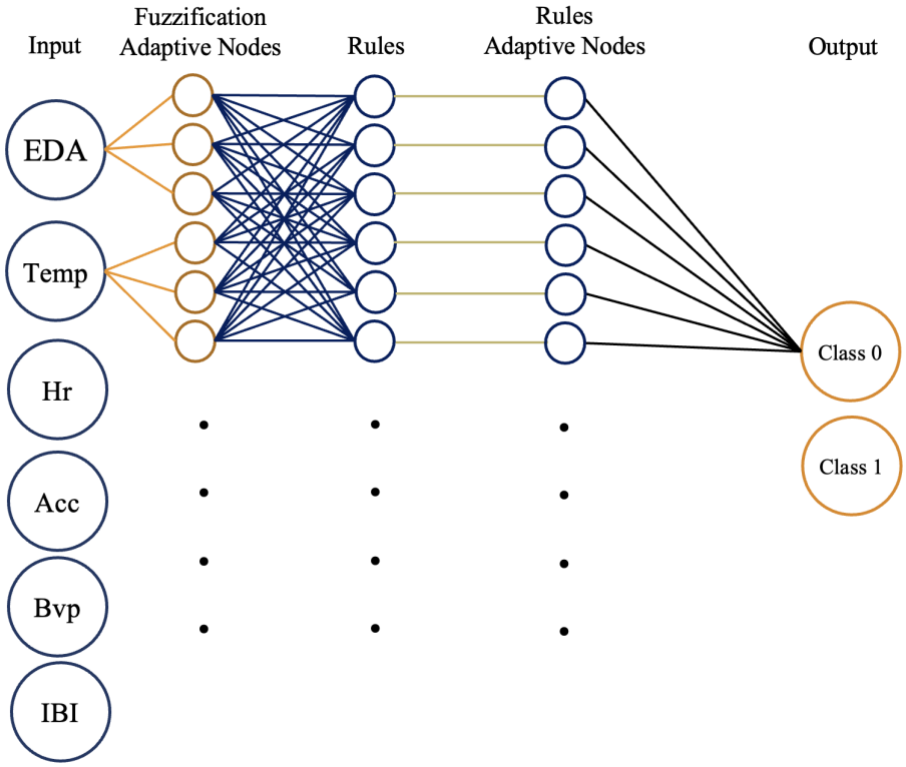

Fig. 35: The ANFIS architecture

6.4.2 Membership Functions (MFs)

The number of membership functions in a fuzzy set varies. As mentioned earlier, a membership function (MF) is a curve that defines in what way each point in a fuzzy set is mapped to a membership value, demonstrating specific linguistic meaning (e.g., small, medium, and large). A higher total number of MFs leads to a higher computational load as it determines the number of rules and adds to the interconnection between nodes. Based on the distribution of each feature in

this research, a reasonable number of MFs are assigned. Ultimately, the optimum number of MFs for each feature can be finalized using hyperparameter tuning and changing the parameter to find the optimum solution.

Table 5 demonstrates the assigned number of membership functions, indexes, and linguistic meanings that are fuzzy values for variables. Variables in this method are the collected physiological signals (i.e., EDA, Temp, Hr, Acc, Bvp, IBI), activity label and reaction. Due to the nature of the accelerometer and the high level of sudden drops and spikes in this variable, the calculated delta introduced ambiguity. It caused the overall low performance of the method. Therefore, one membership function was assigned to this input to minimize the impact on overall performance.

Table 5 Assigned number of membership functions, indexes, and values (linguistic meaning)

Variable	Number of MFs	MF indexes	Fuzzy Values
1.EDA	2	1,2	Small, Large
2.Temp	2	3,4	Small, Large
3.Hr	2	5,6	Small, Large
4.Acc	1	7	
5.Bvp	2	8,9	Small, Large
6.IBI	1	10	
Label	3	11-13	Relax, Physical, Mental Activities

The membership function for all variables is selected to be a bell shape function Eq. (21) presents the mathematical equation, where s, mu, and c are called premise parameters, which are defined in Table 6.

$$\mu_A(x) = \frac{1}{1+\left|\frac{x-mu_i}{s_i}\right|^{2c}} \qquad \text{Eq. (21)}$$

Table 6 Premise and Consequent parameter and their definition for training

Parameters		Definition
Premise	s	Center value of the standard deviation.
	mu	Variation of the mean.
	c	Values of the exponent
Consequent	A	Coefficients matrix containing all consequent parameters

6.4.3 Output Data

Two classes were considered to describe the anomalies' degree of existence: non-anomalous and anomalous states. In this method, output labels for each sample are calculated based on the calculated reconstruction error in section 5.5.

6.4.4 System Specification

Of the total samples, 70% were allocated for training and 30% for testing. That leads to 10386 training samples and 2598 test samples. There are six input variables and two output classes (classes 0 and 1). The proposed ANFIS layout, previously discussed in Table 5, is [2,2,2,1,2,1,3], which indicates the partitioning style. ANFIS layout shows the number of membership functions for each input variable. For instance, two membership functions for the first input variable.

Each membership function in the first layer contains premise functions Eq. (13), which result in an equal number of premise functions. The total number of premise functions is equal to the sum of all numbers in the ANFIS layout (i.e., 13). The number of rules in the network can be calculated

by the product of numbers in the ANFIS layout, which equals the number of interconnections between nodes in the first and second layers (i.e., 48 rules). The number of consequent functions is similar to the number of rules in the network. The proposed ANFIS system specification is presented in Table 7.

Table 7: System specification

System parameters	Specified value
Number of samples	12984
Number of inputs	7
Number of outputs	2
Classes	[0. 1.]
Number of training samples	10386
Number of test samples	2598
ANFIS layout	[2, 2, 2, 1, 2, 1, 3]
Number of premise functions	13
Number of consequent functions	48
Number of model parameter	807

6.4.5 Particle Swarm Optimization (PSO)

Selecting the parameters of the fuzzy neural networks is particularly critical to the ANFIS's success. Primarily, the accuracy and efficiency of the network's learning algorithms determine their success. Gradient-based algorithms are the most common learning algorithm, in which, despite the learning power, it uses the local search technique and tends to result in the local optimum. A particle Swarm Optimization algorithm (PSO) with robust search and optimization ability can help identify optimal or near-optimal solutions in large search spaces. The movement of the Birds Group inspired PSO, and so far, it has been used successfully in many cases (Li et al.,

2011; Y. Zhang & Wu, 2011). PSO is also a good tool for development and application in multi-objective optimization problems due to the high convergence rate commonly found in optimization problems (Hassan et al., 2005).

Solving the problem of local optimum traps related to gradient-based algorithms has interested scholars. One solution is employing collective intelligence algorithms such as PSO algorithms for the training phase. The pseudocode by (Nalluri et al., 2017) for PSO is presented below.

For each particle
    Initialize position and velocity randomly
End
t=1
Do
    For each particle
        Calculate fitness function
        If fitness value > p_{best}, Then
            Set current fitness value as p_{best}
    End
    Update particle with best fitness value as g_{best}
    For each particle
        Calculate new velocity using: $v_i(t) = c_1 r_1 (p_{best}(t) - x_i(t)) + c_2 r_2 (g_{best}(t) - x_i(t))$
        Update position using equation
    End
    t = t+1
While (t < maximum iterations)
Post-process the result.

Researchers have used optimization algorithms to improve the performance of neural network models (Ajbar et al., 2021; Wang et al., 2017; Yu et al., 2008). The ANFIS structure, a neural network-based structure, consists of many nodes in different interconnected layers. The premise and consequent parameters in the nodes determine the ANFIS's performance. PSO was implemented to optimize the proposed ANFIS structure for improved computational performance and to avoid being trapped in local optimums. The flowchart in Fig. 36 presents the process of

using PSO in the proposed ANFIS. It includes the initial iteration for initializing the values and then recurrent iteration to train the mode.

The first iteration, presented on the left side of Fig. 36, sets the basic structures of the system and starts by initializing the dataset information. Then, it follows by setting the PSO initial algorithm parameters (e.g., the number of iterations, population of the particles, coefficients of PSO algorithm, optimization threshold). In the PSO algorithm, the particle's next position is according to three components:

1. A particle velocity,
2. Toward the particle's best performance, and
3. Towards the best performance of all particles.

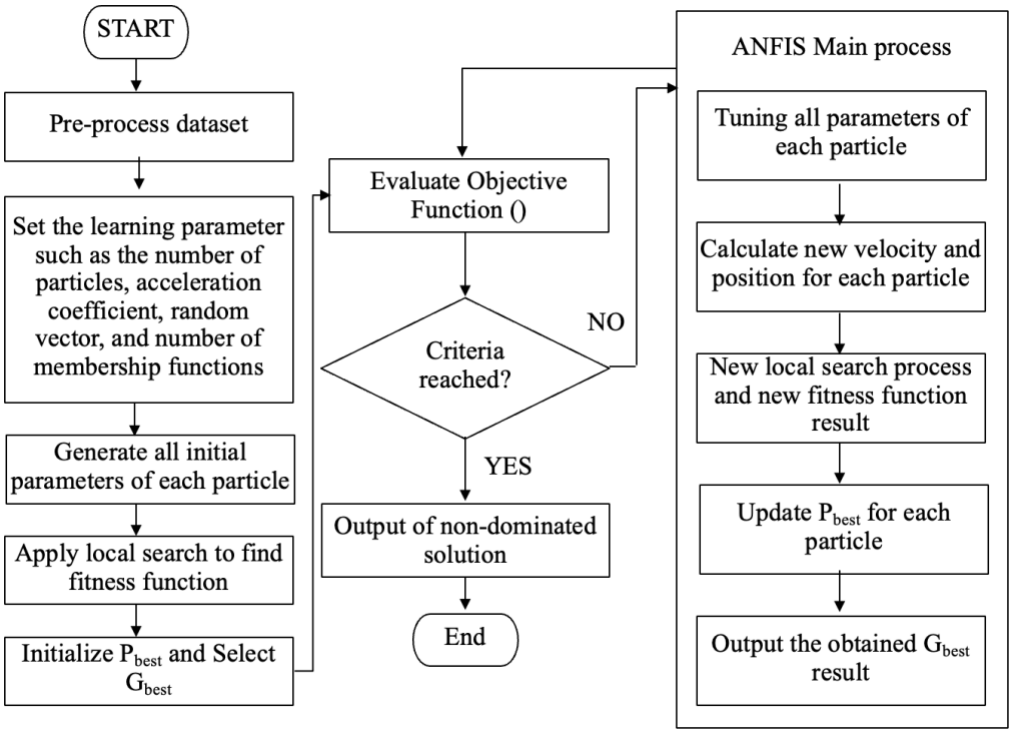

Fig. 36: ANFIS-PSO flowchart

After randomly generating all the initial parameters, the first fitted model is calculated along with memorizing the P_{best} (i.e., local best position) and the G_{best} (i.e., global best position) among all the particles. The results are evaluated using the objective function (i.e., mean-square error MSE). Then, the ANFIS process follows the flowchart in Fig. 36 until the system criteria are reached (i.e., the parameter not improving further in more iterations). Then, the accuracy of the models is checked with statistical indicators. The ANFIS Jupyter Notebook (Eskandar, 2022a) presents the python code, including all calculations and outputs for this method. The source code for ANFIS calculation has been developed by Gilardi, (2020).

The equation of the motion for particles is presented in Eq. (22), in which, x_d is a particle next displacement, υ_d is the particle velocity, p_d is the best performance of the particle, and g_d is the best performance of the global best particle (Clerc, 2010).

$$\begin{cases} \upsilon_d \leftarrow c_1\upsilon_d + c_2(p_d - x_d) + c_3(g_d - x_d) \\ x_d \leftarrow x_d + \upsilon_d \end{cases} \quad \text{Eq. (22)}$$

Traditionally the components of the particle's next movement have been calculated using c_1, c_2, c_3 coefficients. Alternatively, these can be replaced with a confidence coefficient, which makes a linear weighting of three vectors using Eq. (23). Each particle has been given a confidence coefficient for better convergence and robustness.

$$\begin{cases} c_1 = \frac{1}{\varphi - 1 + \sqrt{\varphi^2 - 2\varphi}} \\ c_{max} = \varphi c_1 \end{cases} \quad \text{Eq. (23)}$$

PSO will run simultaneously in the search space to find the optimum value for two objectives: (1) searching for optimal premise/consequent parameters and (2) minimizing the cost function. PSO

defines the search space by setting lower and upper boundaries and searches for the optimal premise and consequent parameters while increasing the system's accuracy.

For finding both presented objectives, 200 particles are assigned to PSO to distribute in the search domain and find the optimum value for all variables mentioned in Table 5. As discussed earlier, the general bell shape function was chosen for MFs. A maximum of 500 iterations are considered to reach the optimum solution before reaching the maximum. The acceleration coefficient presented in Eq. (23) is set to 2.05, looking at similar literature (Qasem & Shamsuddin, 2009). Random vectors r_1 and r_2 are assigned to each particle to define the upper and lower boundaries initial search area. The initial parameter setting of particle swarm optimization (PSO) is presented in Table 8.

6.4.6 Accuracy

The overall system goal is to minimize the cost function, which is classification accuracy. The system's accuracy is measured based on the percentage of correct classifications. The Root Mean Square Error (RMSE) using **Error! Reference source not found.**), or other error aggregation measurements. In classification problems, TP, FP, TN, and FN are designed as below:

- True Positive (TP) is the number of samples correctly identified
- False Positive (FP) is the number of samples incorrectly identified
- True Negative (TN) is the number of samples correctly rejected
- False Negative (FN) is the number of samples incorrectly dismissed.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - t_i)^2}{N}} \qquad \text{Eq. (24)}$$

Table 8 Specification of initial parameter setting

PSO parameter	Value
Number of particles	200
Number of values in a fuzzy set	Different for each variable (Table 5)
Type of membership function	Bell-shape function
Number of iterations	500
Obj. function 1	Find premise/consequent parameters
Obj. function 2	Minimize RMSE to increase accuracy
Acceleration coefficient	φ=2.05
Random vectors r_1 and r_2	Random

As was mentioned earlier, the database is divided into training and testing portions using a 70/30 ratio. Both training and testing databases contain equal samples for each class to remove the biases toward a class that holds the majority. The training set has 10386 samples for two classes (i.e., non-anomalous and anomalous) and 2598 test samples. In training and testing, an equal number of samples is considered for each class to eliminate bias.

6.5 Experimental Result and Analysis

In this section, the results and accuracy of the proposed method are presented. The algorithms, calculations, and results, using Python language programming, are shown in ANFIS Jupyter Notebook (Eskandar, 2022a). An accuracy of 76% was reached, which is considered a reasonable value compared to similar problems (i.e., multi-objective ANFIS for classification (Palupi et al.,

2016)). Indeed, finding the best membership function and defining rules based on satisfactory accuracy is not easy.

Critical values of True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) were calculated based on the confusion matrix (Fig.37), shown in Table 9. False negatives for non-anomalous points are higher, leading to low precision for anomalous points.

Table 9: Calculated TP, FP, FN, and TN for class 0 (Non-Anomalous) and class 1 (Anomalous)

Class		TP	FP	FN	TN
0	Non-Anomalous	899	227	400	1072
1	Anomalous	1072	400	227	899

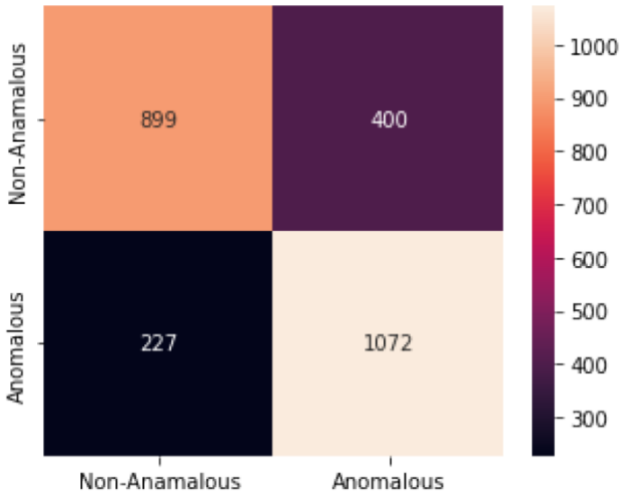

Fig. 37: Confusion Matrix for the test dataset

Classification measurements of the ANFIS method are presented in Table 10 which, similar to the confusion matrix, offers high precision, recall, and F1-score for the Anomalous and the Non-Anomalous classes.

Table 10 similar to what was explained in the previous chapter, section 5.5.2, presents precision Eq. (9), which response to how many predicted anomalies were truly anomalous out of all detected anomalies. Recall Eq. (10), aka sensitivity, response to how many predicted anomalies were truly anomalies out of all actual existing anomalies. F1-score Eq. (11) considers both precision and recall. The macro average considers all classes equally, irrespective of the number of support values. The weighted-averaged is calculated by taking the mean of F1-scores for classes while considering each class's support number. Accuracy Eq. (12), or micro F1, is a global F1-score average.

Table 10: ANFIS accuracy

	Precision	Recall	F1-score	Support
Non- Anomalous	0.80	0.69	0.74	1299
Anomalous	0.73	0.83	0.77	1299
Accuracy			0.76	2598
Macro avg	0.76	0.76	0.76	2598
Weighted avg	0.76	0.76	0.76	2598

The ANFIS method has a 0.8 precision for non-anomalies and 0.73 for anomalies. The method can recall 0.69 non-anomalies and 0.83 anomalies. F1-Score for anomalies is 0.77 and 0.74 for non-anomalies. Accuracy, macro averaged, and weighted averaged F1-score is 0.76. As mentioned in previous sections, the ANFIS method was trained and tested using the same number of samples for both classes due to the algorithm's sensitivity to imbalanced distribution in classification.

The method performance is balanced with 0.74 and 0.77 F1-score for non-anomalies and anomalies. Applying the ANFIS method leads to experiencing low false alarms, and it also fails to

detect some of the anomalies. The overall performance of the ANFIS method is better for the anomalous class with the 0.77 F1-score compared to the 0.52 value by the first method. More comparisons are presented in the final chapter.

Table 11 presents the result of ANFIS sensitivity and specificity for each class that shows low sensitivity for Non-Anomalous, similar to recall, is the accuracy for predicting classes (i.e., the true number of positives per total actual positives). Also, there is comparably low specificity for the Anomalous class. Specificity is the accuracy for predicting non-events (i.e., the true negatives per total actual negatives of a classifier).

Table 11 Sensitivity and specificity for classes

	Class	Specificity	Sensitivity
0	Non- Anomalous	0.82525	0.692071
1	Anomalous	0.692071	0.82525

Table 12 contains the calculated premise parameter during the training session that leads to the optimum value for objectives. As presented in Table 6, the bell-shaped membership function parameters are mu, s, and c, which correspond to the mean, standard deviation, and exponent value of the general bell-shaped function Eq. (21). ANFIS method's two objectives were to search for premise parameters that define membership functions and consequent petameters that define rules. Based on the values presented in Table 12, membership functions for every seven variables are plotted in Fig. 38 to Fig. 44 . Linguistic meaning of membership functions is shown in Table 5.

Table 12: Premise membership parameter

	Inputs	No. of MFs	Membership Parameter								
			1			2					
			c	s	mu	c	s	mu			
1	EDA	2	1.249	16.558	-0.255	1.044	16.495	32.775			
2	Temp	2	1.536	25.610	-0.441	1.007	25.535	51.276			
3	Hr	2	2.239	19.863	0.387	1.200	19.950	36.195			
4	Acc	1	1.748	0.429	6.724						
5	Bvp	2	1.633	7.756	0.077	1.416	7.715	11.937			
6	IBI	1	2.807	0.419	1.651						
7	label	3	1.738	0.825	-0.363	1.177	0.673	0.689	2.042	0.946	2.102

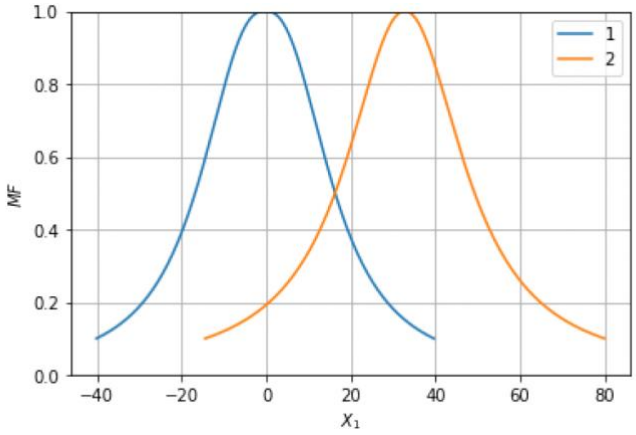

Fig. 38: Calculated membership function for EDA

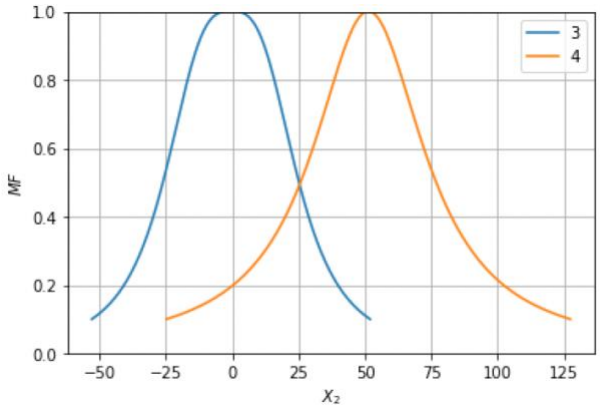

Fig. 39: Calculated membership function for Temp

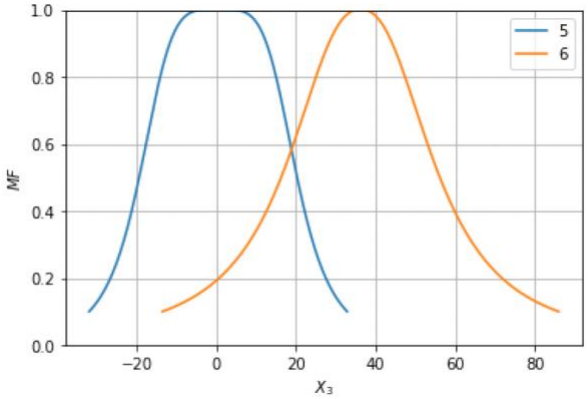

Fig. 40: Calculated membership function for Hr

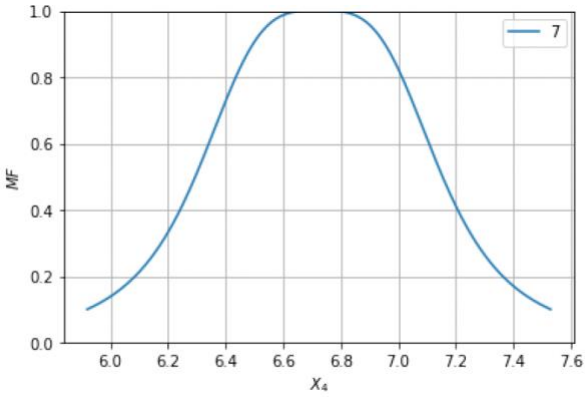

Fig. 41: Calculated membership function for Acc

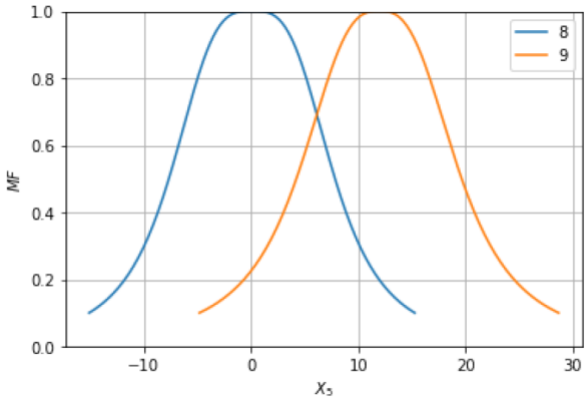

Fig. 42: Calculated membership function for Bvp

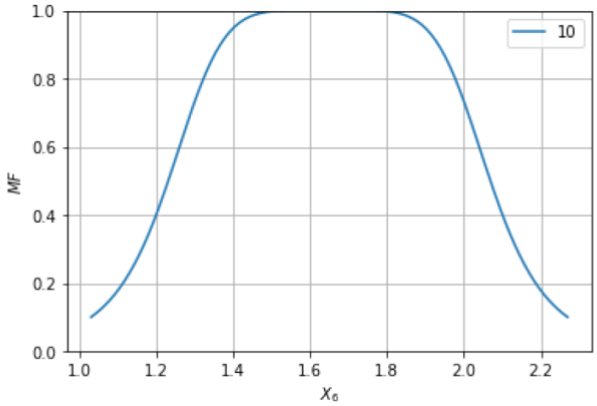

Fig. 43: Calculated membership function for IBI

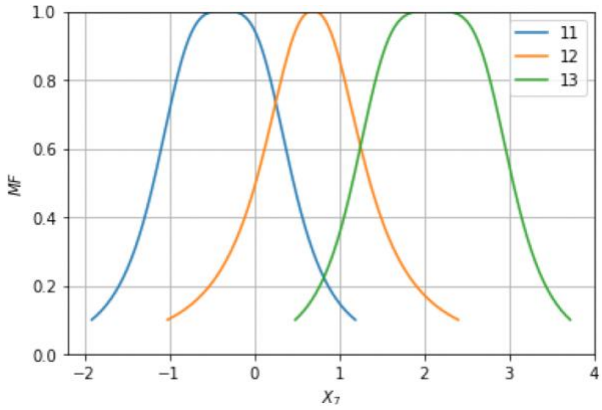

Fig. 44: Calculated membership function for Labels

Due to the vast volume of interconnection between membership functions, the calculated consequent parameters matrix's length is 96, which presents consequent parameters for all interconnection between nodes with membership functions. The matrix can be found in ANFIS Jupyter Notebook (Eskandar, 2022a).

6.6 Conclusion

An adaptive neuro-fuzzy system (ANFIS) has been proposed to classify and partition data into two classes representing anomalies' degrees of existence. This method assigns optimum membership functions to each partition and allocates the rules. Adaptive layers in this network improve the parameter over iterations using an optimization algorithm efficient for the vast search area. Particle swarm optimization (PSO) searches the parameters by assigning particles and moving each particle toward the global optimum. The root-mean-square error is used to calculate the training error at each iteration and is backpropagated to improve membership functions and rules in adaptive layers. The ANFIS method is adaptive and selected due to the absence of reasoning about the inherently vague concept of abnormal psychological states. Typical FIS requires defining membership functions for all values and rules among all values. Due to the absence of a calculated equation for each physiological signal and knowing the relationship between values, an adaptive method is necessary.

The experimental results demonstrated the effectiveness of the proposed approach for identifying Anomalous and Non-Anomalous states. The method performance of the proposed method is balanced with 0.74 and 0.77 F1-score for non-anomalies and anomalies class. This method leads to experiencing low false alarms (0.73 precision for anomalies class). The overall performance of

the ANFIS method is better for the anomalous class with the 0.77 F1-score compared to the 0.52 value by the first method. More comparisons are presented in the final chapter.

More importantly, it does not require any user-specified parameter for either setting rules or defining membership functions. This method addresses the shortcomings of the previous approach by not relying on the threshold for determining anomalies. However, it resulted in comparably lower accuracy than the previous method. The next chapter discusses each approach with its performance and compares the outcomes.

Chapter 7 Conclusions

7.1 Chapter Overview

This research was inspired by identifying the strong association of stress with unsafe behaviour in construction. The study investigates the physiological features associated with stress and offers enhanced stress detection methods resulting in contributions discussed further in this chapter. This chapter first summarizes the research study presented in this dissertation and discusses the main contributions to the body of knowledge in data analytics for safety. The chapter follows with discussions on limitations and potential future work for the field experiment design and data collection process, as well as the two stress detection methods based on unsupervised deep learning and the Adaptive Neuro-Fuzzy Inference System. A discussion on the comparison of the two methods follows. The chapter concludes with concluding remarks.

7.2 Research Summary

A comprehensive literature review revealed that current stress detection methods work based on limited physiological signals without considering the subjects' basic information or situational data. Many field experiments for stress detection research are designed with inadequate stress-inducing approaches during data collection. Also, among the previous studies, there is no applied anomaly detection-based method in the stress detection domain. These identified gaps in research

motivated this study to develop stress detection methods that involve anomaly detection and are trained using the collected dataset from the designed field experiments. The conducted data collection offers diverse records of data collected through unobtrusive methods while using stress-inducing approaches.

This study presented two anomaly detection methods. For the first method, an unsupervised deep learning-based method was used for detecting workers' stress and overload based on the acquired data. This method performs based on three categories of acquired data. The proposed method includes unsupervised learning architecture formulated over an input size of (5,14,14,3). Input is the correlation of 14 features stacked together for five consecutive timesteps (the previous 1.25 seconds) and three different window sizes.

The method performs based on reconstructing the same input (i.e., pairwise correlation). High reconstruction error indicates the method's inability to predict, therefore flagging an anomaly. The proposed anomaly detection method supports identifying contributing features for the detected anomalies. Interpretation is possible by analyzing the residual matrix and looking into error values for correlation between every two features. The methods' limitations and shortcomings, discussed in Chapter 6, motivated further investigation and development of the Fuzzy Logic method.

Anomaly detection using an Adaptive Neuro-Fuzzy Inference System (ANFIS) was designed and developed to classify and partition data into two classes representing anomalies' degrees of existence. This method performs based on six physiological signals and labels of the performed activities for five consecutive timesteps in the previous 1.25 seconds, which includes input size equal to (5, 7). ANFIS method assigns optimum membership functions to each partition and allocates the rules. Particle swarm optimization (PSO) is used to search the parameters. The root-

mean-square error is calculated at each iteration and backpropagated to improve membership functions. The experimental results demonstrated the effectiveness of the proposed approach, especially for identifying Anomalous and Non-Anomalous cases.

7.3 Research Contributions

This research pursued three objectives presented in section 1.3. The first objective was to design and implement a field experiment to collect a range of data, including physiological and contextual data while inducing stress. As demonstrated in Chapter 4, section 4.2, a field experiment was conducted with 18 participants performing a pre-defined sequence of activities triggering mental and physical stress. A wristband biosensor is used to collect physiological signals while engaging in activities. The database of contextual and physiological data was created to be utilized for training and testing stress detection methods.

The second objective was to design, develop, and evaluate anomaly detection method(s) for stress and overload detection. As given in Chapter 5 and Chapter 6, two methods were developed in response to the second objective. The third objective was to align the model with the long-term goals and future application in the construction industry as a potential human in the loop cyber-physical system. This was partially achieved through the experiment design stage. Participants were provided with PPE equipment, and unobtrusive wearable sensors were selected. Collecting contextual data on the work condition and tasks in the future and using it in the model design will align the stress detection model to different workplaces.

The overall outcome of this study is an advanced stress detection toolkit for improving safety in workplaces, with potential applications to other areas such as education, driving, and long-term

care. The improvements in the stress detection approach were achieved using the power of data, AI, and ML. AI/ML-enhanced data analytics was the enabler of this work, contributing to the body of knowledge in safety research. This study proved effective in developing an anomaly detection method using two methods: (1) Unsupervised DL and (2) ANFIS.

In terms of visualization, interpretation, and application, the first method provides reconstruction error matrices that allow us to analyze error values, which is the correlation difference between every two features. Higher intensity of error values regards as the abnormality from commonly recognized patterns. On the other hand, the second method is easier to utilize by a non-specialist in the machine learning domain. This method presents the membership functions and fuzzy rules, which are easier to understand and deploy.

7.4 Limitations and Future Work

7.4.1 Experiment Design and Data Collection

The proposed stress detection methods heavily rely on data. The development and validation of the methods presented in this study are influenced by the quality and limitations of the data collected through the field experiments. It is important to note that the field experiment was conducted in controlled circumstances and therefore lacks the actual work setting data. In addition, only a limited number of activities and participants were considered in the experiment. For better detection performance for specific workplaces, it is recommended that the method be trained on a broader range of subjects and activities for future work, with the training/testing data collected in an actual dynamic situation in the targeted workplace.

There are limitations to effectively collecting some non-invasive physiological signals from construction workers while involved in physical activities. For instance, utilizing precise and effective ways of collecting data that show the impact of stress and overload, such as EEG signals, is not practical. EEG signals are sensitive to intrinsic artifacts like eye blinking in addition to extrinsic artifacts that arise from the outside body (e.g., environmental noises) that bring some limitations in their applications. More contextual information can be recorded to enhance the system further, such as the participant's task, location, and surrounding noise.

Moreover, safety is not related to the isolated act of individuals, and the social interactions resulting from the collaboration and communication among workers in such a dynamic environment will impact human behaviour. Therefore, investigating social factors and their influence on stress and situation management is an important area to be examined in future work to advance safety research. Moreover, researchers' future work includes stress detection model adaptation and calibration in different workplaces such as construction. The reliability of the detected anomalies will be further improved using a decision fusion method to calibrate the anomaly detection method for any different workspace or application context.

7.4.2 Stress Detection Using Unsupervised Deep Learning

The method was examined against two separate test datasets resulting in 98 percent accuracy over the first test set. The labels (i.e., absolute truth) for the first test dataset are tagged on the expert's opinion by studying the raw collected time series. For future work, including different expert inputs and specialists (e.g., Cardiologists) is recommended to improve the absolute truth in calculating the accuracy.

The second test dataset was arranged without following the original sequence of activities in order to prove the method's independence from the sequence of activities in the designed experiment. While recording data, the times of intentionally added extra mental stress/overload were logged as labels. So, the second test dataset size includes tagged timesteps based on real-time observations. The method showed 90 percent accuracy against the second test dataset. This reduced performance can be attributed to the significant difference in individual responses to the same stress stimuli that require future extensive research over customizing anomaly detection models for every individual. It is unrealistic to assume that the outcomes can be compared against absolute truth in a problem space, which is uncertain by nature. Therefore, the stress detection method discussed in this chapter suffers from a lack of availability of absolute truth. Anomalies in this method are identified when the reconstruction error is higher than the defined threshold (90 percentile of the reconstruction error for each activity).

Defining a higher threshold will result in high false negatives, while a lower value leads to high false positives. So, setting thresholds and over-reliance on absolute truth are the shortcomings of the DL-based anomaly detection method. To overcome the above-mentioned challenges and shortcomings, further research and investigation resulted in the designing and development of an adaptive Neuro-Fuzzy Inference System for anomaly detection.

7.4.3 Stress Detection Using Adaptive Neuro-Fuzzy Inference System

The second method addresses the shortcomings of the previous approach by not relying on the threshold for determining anomalies. Moreover, the proposed ANFIS, unlike common FIS, does not require any user-specified parameter for either setting rules or defining membership functions.

This method is faster in nature and offers a transparent distribution of input data through a Membership Function (MF) that provides a straightforward approach to users. However, the ANFIS method resulted in considerably lower accuracy than the deep learning method. This study considered a general bell shape MF for all input variables. To improve the accuracy, the impact of different forms of membership functions can be added to pick the best matching form automatically. General bell shape MF was considered for all input variables, and diverse MFs can be assigned to each variable. Moreover, increasing the flexibility of MFs and rules can be regarded as considering wider combinations of options to be searched by an optimization algorithm.

Another critical issue in stress detection methods is the natural differences in individuals responding to stress-inducing events or conditions. Further studies can address this limitation to customize and calibrate anomaly detection methods for every individual.

7.5 Methods' Comparison and Discussion

7.5.1 Methods Efficiency and Training Cost

Regarding the efficiency and training cost, the unsupervised deep learning-based method has 144 times larger parameters to train compared with the adaptive neuro-fuzzy inference system method. The first method contains a total of 116,028 parameters to be trained; however, there are only 807 parameters for the second method. The higher number of parameters indicates a higher training cost for the first method. Above all, the first method requires more samples for the training process, which undoubtedly results in a higher data collection and management cost. Both methods' processing time was calculated using Intel(R) Xeon(R) CPU @ 2.30GHz CPU. The first method

reached the stopping criteria in 1d 5h 16min 26s, which is 16 times slower than 1h 49min 49s for the second method.

Table 13: Unsupervised DL-based and ANFIS anomaly detection efficiency and training cost

Methods	Parameters	Processing Time
1-Unsupervised DL-Based	116028	1d 5h 16min 26s
2-ANFIS	807	1h 49min 49s

7.5.2 Methods' Performance Metrics

Both methods' performance metrics are organized in Table 14, which presents the performance of the first method (i.e., 1- Unsupervised DL-Based) next to the second method (i.e., 2-ANFIS). The first method has a precision of 1.00 for non-anomalies data, while the second method offers a precision of 0.8 for the same class of data. For the anomalies, the second method provides a precision of 0.73, which outperforms the first method with a precision of 0.35. The first method recall is 0.98 and 1 for non-anomalies and anomalies, higher than 0.69 and 0.83 for the same classes in the second method. The second method's F1-score for anomalies is 0.77, higher than the 0.52 of the first method. However, the non-anomalies F1-Score is 0.99 for the first method, which outperforms the 0.74 for the second method.

There is a considerable difference between the two methods' accuracy. Accuracy as a global F1-score average is the percentage of correct predictions over the total number of samples. The first method resulted in 0.98 accuracies, higher than the second method with 0.76 accuracies. Accuracy, macro average, and the weighted average for the second method are equal (i.e., 0.76) since the

second method was trained and tested with the same number of samples for each class. As explained in Chapter 6, the second method holds the same number of samples for both classes due to the algorithm's sensitivity to unbalanced distribution in classification.

The macro averaged F1-score considers all classes equally, irrespective of the class weight, which in imbalanced class distribution, there are different weights to different classes. The weighted averaged F1-score is calculated by taking the mean of all F1-scores for classes while considering the class's weight (i.e., the proportion of samples for each class). The first method, with 0.99 weighted averaged F1-score, performs better than the second method, with 0.76 weighted averaged F1-score. Overall, the second method performs better for the anomalous class with the 0.77 F1-score compared to the 0.52 value by the first method.

Table 14 Unsupervised DL-based and ANFIS anomaly detection accuracy

	Anomaly Detection Method					
	1-Unsupervised DL-Based			2-ANFIS		
Classes	Precision	Recall	F1-score	Precision	Recall	F1-score
Non- Anomalous	1.00	0.98	0.99	0.80	0.69	0.74
Anomalous	0.35	1.00	0.52	0.73	0.83	0.77
Accuracy			0.98			0.76
Macro avg	0.68	0.99	0.76	0.76	0.76	0.76
Weighted avg	0.99	0.98	0.99	0.76	0.76	0.76

Based on the defined metrics, the first method can predict all anomalies. However, it is detecting more anomalies than exist due to low precision. Therefore, the first method is suitable for highly sensitive circumstances where the false alarm is accepted for a higher level of safety. However, the second method is more balanced regarding detecting anomalies, with 0.83 recall and 0.73

precision. Applying the second method leads to experiencing fewer false alarms while not detecting some of the anomalies.

7.6 Conclusions

This research demonstrated improvements in the stress detection approach using the power of data, AI, and Machine Learning. AI/ML-enhanced data analytics was the enabler of this work, contributing to the body of knowledge in safety research. However, as discussed, the nature of the stress phenomena, the construction environment, and the limitations of data collection methods constrained this work.

Despite the above-mentioned fundamental limitations, this study proved effective in developing an anomaly detection method using two methods: (1) unsupervised deep learning-based and (2) adaptive neuro-fuzzy inference system with 98% and 76% accuracy.

In conclusion, the second method performs better for the anomalous class than the first method. Also, the second method is more balanced regarding detecting anomalies and non-anomalies, which leads to lower and more manageable false alarms. However, the first method is suitable for highly sensitive circumstances where the false alarm is accepted.

REFERENCES

Abadi, M. , Agarwal, A. , Barham, P. , Brevdo, E. , Chen, Z. , Citro, C. , Corrado, G. S. , Davis, A. , Dean, J. , Devin, M. , Ghemawat, S. , Goodfellow, I. , Harp, A. , Irving, G. , Isard, M. , Jozefowicz, R. , Jia, Y. , Kaiser, L. , Kudlur, M. , … Zheng, X. (2015). *TensorFlow, Large-scale machine learning on heterogeneous systems* (2.8.3). https://doi.org/https://doi.org/10.5281/zenodo.4724125

Abdelhamid, T. S., & Everett, J. G. (2000a). Identifying root causes of construction accidents. *Journal of Construction Engineering and Management*, *126*(1), 52–60. https://doi.org/10.1061/(ASCE)0733-9364(2000)126:1(52)

Abdelhamid, T. S., & Everett, J. G. (2000c). Ironworkers: Physiological demands during construction work. *Proceedings of Construction Congress VI: Building Together for a Better Tomorrow in an Increasingly Complex World*, *278*(October), 631–639. https://doi.org/10.1061/40475(278)68

Abdelhamid, T. S., & Everett, J. G. (2002). Physiological demands during construction work. *Journal of Construction Engineering and Management*, *128*(5), 427–437. https://doi.org/10.1061/(ASCE)0733-9364(2002)128:5(427)

Abuwarda, Z., Hegazy, T., Oetomo, A., & Morita, P. P. (2023). Using Wearables to Monitor and Mitigate Workers' Fatigue. *Lecture Notes in Civil Engineering*, *247*, 587–597. https://doi.org/10.1007/978-981-19-0968-9_47/COVER

Adams, K. N., Szumowski, J. D., & Ramakrishnan, L. (2014). Verapamil, and its metabolite norverapamil, inhibit macrophage-induced, bacterial efflux pump-mediated tolerance to

multiple anti-tubercular drugs. *Journal of Infectious Diseases*, *210*(3), 456–466. https://doi.org/10.1093/infdis/jiu095

Airij, A. G., Sudirman, R., & Sheikh, U. U. (2018). GSM and GPS based real-time remote physiological signals monitoring and stress levels classification. *2nd International Conference on BioSignal Analysis, Processing and Systems, ICBAPS 2018*, 130–135. https://doi.org/10.1109/ICBAPS.2018.8527406

Ajbar, W., Parrales, A., Cruz-Jacobo, U., Conde-Gutiérrez, R. A., Bassam, A., Jaramillo, O. A., & Hernández, J. A. (2021). The multivariable inverse artificial neural network combined with GA and PSO to improve the performance of solar parabolic trough collector. *Applied Thermal Engineering*, *189*, 116651. https://doi.org/10.1016/J.APPLTHERMALENG.2021.116651

Alarifi, I. M., Nguyen, H. M., Bakhtiyari, A. N., & Asadi, A. (2016). *materials Feasibility of ANFIS-PSO and ANFIS-GA Models in Predicting Thermophysical Properties of Al 2 O 3-MWCNT/Oil Hybrid Nanofluid*. https://doi.org/10.3390/ma12213628

Anumba, C. J., Akanmu, A., & Messner, J. (2010). Towards a cyber-physical systems approach to construction. *Construction Research Congress 2010: Innovation for Reshaping Construction Practice - Proceedings of the 2010 Construction Research Congress*, 528–537. https://doi.org/10.1061/41109(373)53

Anusha, A. S., Sukumaran, P., Sarveswaran, V., Surees Kumar, S., Shyam, A., Akl, T. J., Preejith, S. P., & Sivaprakasam, M. (2020). Electrodermal Activity Based Pre-surgery Stress Detection Using a Wrist Wearable. *IEEE Journal of Biomedical and Health Informatics*, *24*(1), 92–100. https://doi.org/10.1109/JBHI.2019.2893222

AWCBC/ACATC. (2020). *Association of Workers' Compensation Boards of Canada. 2020 National Work Injury, Disease and Fatality Statistics*. http://awcbc.org/?page_id=14#fatalities

Bali, A., & Jaggi, A. S. (2015). Clinical experimental stress studies: methods and assessment. *Reviews in the Neurosciences*, *26*(5), 555–579. https://doi.org/10.1515/REVNEURO-2015-0004

Bari, D. S., Aldosky, H. Y. Y., Tronstad, C., & Martinsen, Ø. G. (2021). The correlations among the skin conductance features responding to physiological stress stimuli. *Skin Research and Technology*, *27*(4), 582–588. https://doi.org/10.1111/SRT.12989

Bedny, G., & Meister, D. (1999). Theory of Activity and Situation Awareness. *International Journal of Cognitive Ergonomics*, *3*(1), 63–72. https://doi.org/10.1207/s15327566ijce0301_5

Behzadan, A. H., Dong, S., & Kamat, V. R. (2015). Augmented reality visualization: A review of civil infrastructure system applications. *Advanced Engineering Informatics*, *29*(2), 252–267. https://doi.org/10.1016/J.AEI.2015.03.005

Betti, S., Lova, R. M., Rovini, E., Acerbi, G., Santarelli, L., Cabiati, M., Ry, S. Del, & Cavallo, F. (2018). Evaluation of an integrated system of wearable physiological sensors for stress monitoring in working environments by using biological markers. *IEEE Transactions on Biomedical Engineering*, *65*(8), 1748–1758. https://doi.org/10.1109/TBME.2017.2764507

Biondi, F. N., Cacanindin, A., Douglas, C., & Cort, J. (2021). Overloaded and at Work: Investigating the Effect of Cognitive Workload on Assembly Task Performance. *Human Factors*, *63*(5), 813–820. https://doi.org/10.1177/0018720820929928

Birjandtalab, J., Cogan, D., Pouyan, M. B., & Nourani, M. (2016). A non-EEG biosignals dataset for assessment and visualization of neurological status. *IEEE Workshop on Signal Processing Systems, SiPS: Design and Implementation*, 110–114. https://doi.org/10.1109/SiPS.2016.27

Bonomi, F., Milito, R., Zhu, J., & Addepalli, S. (2012). Fog computing and its role in the internet of things. *MCC'12 - Proceedings of the 1st ACM Mobile Cloud Computing Workshop*, 13–15. https://doi.org/10.1145/2342509.2342513

Brisswalter, J., Collardeau, M., & René, A. (2012). Effects of Acute Physical Exercise Characteristics on Cognitive Performance. *Sports Medicine 2002 32:9*, *32*(9), 555–566. https://doi.org/10.2165/00007256-200232090-00002

Calibo, T. K., Blanco, J. A., & Firebaugh, S. L. (2013). Cognitive stress recognition. *Conference Record - IEEE Instrumentation and Measurement Technology Conference*, 1471–1475. https://doi.org/10.1109/I2MTC.2013.6555658

Can, Y. S., Arnrich, B., & Ersoy, C. (2019). Stress detection in daily life scenarios using smart phones and wearable sensors: A survey. *Journal of Biomedical Informatics*, *92*, 103139. https://doi.org/10.1016/J.JBI.2019.103139

Can, Y. S., Chalabianloo, N., Ekiz, D., & Ersoy, C. (2019). Continuous Stress Detection Using Wearable Sensors in Real Life: Algorithmic Programming Contest Case Study. *Sensors 2019, Vol. 19, Page 1849*, *19*(8), 1849. https://doi.org/10.3390/S19081849

Can, Y. S., Chalabianloo, N., Ekiz, D., Fernandez-Alvarez, J., Riva, G., & Ersoy, C. (2020). Personal Stress-Level Clustering and Decision-Level Smoothing to Enhance the Performance of Ambulatory Stress Detection with Smartwatches. *IEEE Access*, *8*, 38146–38163. https://doi.org/10.1109/ACCESS.2020.2975351

Cárdenas, A. A., Amin, S., & Sastry, S. (2008). Secure control: Towards survivable cyber-physical systems. *Proceedings - International Conference on Distributed Computing Systems*, 495–500. https://doi.org/10.1109/ICDCS.Workshops.2008.40

Cárdenas, A. A., Amin, S., Sinopoli, B., Giani, A., Perrig, A., & Sastry, S. (2009). Challenges for Securing Cyber Physical Systems. *Prevention*, submitted} @techreport{LEE:2007:ID2217, author=. http://chess.eecs.berkeley.edu/pubs/416.html

Chen, J., Taylor, J. E., & Comu, S. (2017). Assessing Task Mental Workload in Construction Projects: A Novel Electroencephalography Approach. *Journal of Construction Engineering and Management*, *143*(8), 04017053. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001345

Cheng, T., & Teizer, J. (2013). Real-time resource location data collection and visualization technology for construction safety and activity monitoring applications. *Automation in Construction*, *34*, 3–15. https://doi.org/10.1016/j.autcon.2012.10.017

Cheng, Z., Wang, S., Zhang, P., Wang, S., Liu, X., & Zhu, E. (2021). Improved autoencoder for unsupervised anomaly detection. *International Journal of Intelligent Systems*, *36*(12), 7103–7125. https://doi.org/10.1002/INT.22582

Choudhry, R. M., Fang, D., & Lingard, H. (2009). Measuring safety climate of a construction company. *Journal of Construction Engineering and Management*, *135*(9), 890–899. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000063

Choudhry, R. M., Fang, D., & Mohamed, S. (2007). The nature of safety culture: A survey of the state-of-the-art. *Safety Science*, *45*(10), 993–1012. https://doi.org/10.1016/j.ssci.2006.09.003

Chua, D. K. H., & Goh, Y. M. (2004). Incident causation model for improving feedback of safety knowledge. *Journal of Construction Engineering and Management*, *130*(4), 542–551. https://doi.org/10.1061/(ASCE)0733-9364(2004)130:4(542)

Cisco Systems. (2016). Fog Computing and the Internet of Things: Extend the Cloud to Where the Things Are. *Www.Cisco.Com*, 6. http://www.cisco.com/c/dam/en_us/solutions/trends/iot/docs/computing-overview.pdf

Cooke, N., & Chong, O. (2017). *Human-Centered Automation for Resilience in Acquiring Construction Field Information by Cheng Zhang* (Issue May). ARIZONA STATE UNIVERSITY.

Cooper, C. L. (2005). *Handbook of stress medicine and health*. 410.

Dawood, N., & Mallasi, Z. (2006). Construction workspace planning: Assignment and analysis utilizing 4D visualization technologies. *Computer-Aided Civil and Infrastructure Engineering*, *21*(7), 498–513. https://doi.org/10.1111/j.1467-8667.2006.00454.x

Deary, I. J. (1996). Measuring stress: A guide for health and social scientists. *Journal of Psychosomatic Research*, *41*(2), 186. https://doi.org/10.1016/0022-3999(96)00066-9

Deepak, K., Chandrakala, S., & Mohan, C. K. (2021). Residual spatiotemporal autoencoder for unsupervised video anomaly detection. *Signal, Image and Video Processing*, *15*(1), 215–222. https://doi.org/10.1007/S11760-020-01740-1/FIGURES/6

Depari, A., Flammini, A., & Marioli, D. (2007). Application of an ANFIS algorithm to sensor data processing. *Ieeexplore.Ieee.Org*. https://ieeexplore.ieee.org/abstract/document/4061079/

Dobson, A. (2003). *The Oxford Dictionary of Statistical Terms*.

Dogantekin, E., Dogantekin, A., & Avci, D. (2010). An intelligent diagnosis system for diabetes on linear discriminant analysis and adaptive network based fuzzy inference system: LDA-ANFIS. *Elsevier- Digital Signal*. https://www.sciencedirect.com/science/article/pii/S1051200409001997

Egilmez, B., Poyraz, E., Zhou, W., Memik, G., Dinda, P., & Alshurafa, N. (2017). UStress: Understanding college student subjective stress using wrist-based passive sensing. *2017 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2017*, 673–678. https://doi.org/10.1109/PERCOMW.2017.7917644

Endsley, M. R. (1988a). A construct and its measurement: The functioning and evaluation of pilot situation awareness. *Northrop Technical Report: NOR DOC 88-30*.

Endsley, M. R. (1988b). Situation Awareness Global Assessment Technique (Sagat). *IEEE Proceedings of the National Aerospace and Electronics Conference*, 789–795.

Eskandar, S. (2022a). *Jupyter Notebooks ANFIS.ipynb*. https://github.com/Sahel-Eskandar/PhD_Thesis/blob/main/ANFIS.ipynb

Eskandar, S. (2022b). *Jupyter Notebooks Deep-Learning.ipynb*. https://github.com/Sahel-Eskandar/PhD_Thesis/blob/main/Deep_Learning.ipynb

Eskandar, S. (2022c). *Jupyter Notebooks Pre_Processing.ipynb*. https://github.com/Sahel-Eskandar/PhD_Thesis/blob/main/Pre_Processing.ipynb

Eskandar, S., & Razavi, S. (2020). Using deep learning for assessment of workers' stress and overload. *Proceedings of the 37th International Symposium on Automation and Robotics in Construction, ISARC 2020: From Demonstration to Practical Use - To New Stage of Construction Robot*, 872–877. https://doi.org/10.22260/ISARC2020/0120

Eskandar, S., Wang, J., & Razavi, S. (2019). A review of social, physiological, and cognitive factors affecting construction safety. *Proceedings of the 36th International Symposium on Automation and Robotics in Construction, ISARC 2019*, *Isarc*, 317–323. https://doi.org/10.22260/isarc2019/0043

Fang, D., Zhao, C., & Zhang, M. (2016). A cognitive model of construction workers' unsafe behaviors. *Journal of Construction Engineering and Management*, *142*(9), 04016039. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001118

Fix, E., & Hodges, J. L. (1989). Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. *International Statistical Review / Revue Internationale de Statistique*, *57*(3), 238. https://doi.org/10.2307/1403797

Gedam, S., & Paul, S. (2021). A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques. In *IEEE Access* (Vol. 9, pp. 84045–84066). https://doi.org/10.1109/ACCESS.2021.3085502

Genders, W., Wang, J., Razavi, S., Candidate, P. D., Engineering, C., & Email, C. (2015). Smartphone Construction Safety Awareness System : A Cyber-Physical System Approach. *The 16th International Conference on Computing in Civil and Building Engineering (ICCCBE2016)*, 1697–1704.

Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., & Tsiknakis, M. (2022). Review on Psychological Stress Detection Using Biosignals. *IEEE Transactions on Affective Computing*, *13*(1), 440–460. https://doi.org/10.1109/TAFFC.2019.2927337

Gilardi, G. (2020). *Multivariate Regression and Classification Using an Adaptive Neuro-Fuzzy Inference System (Takagi-Sugeno) and Particle Swarm Optimization*. https://github.com/gabrielegilardi/ANFIS

Goldenhar, L. M., Williams, L. J., & Swanson, N. G. (2003). Modelling relationships between job stressors and injury and near-miss outcomes for construction labourers. *Work and Stress*, *17*(3), 218–240. https://doi.org/10.1080/02678370310001616144

Government of Canada. (2020). *Hard-surface disinfectants and hand sanitizers (COVID-19)*. https://www.canada.ca/en/health-canada/services/drugs-health-products/disinfectants/covid-19/list.html

Griffor, E., Wollman, D., & Greer, C. (2017). Framework for Cyber-Physical Systems: Volume 1, Overview. In *NIST Special Publication* (Vol. 1, Issue June). https://doi.org/10.6028/NIST.SP.1500-201

Gualdi, G., Prati, A., & Cucchiara, R. (2009). Covariance descriptors on moving regions for human detection in very complex outdoor scenes. *2009 3rd ACM/IEEE International Conference on Distributed Smart Cameras, ICDSC 2009*. https://doi.org/10.1109/ICDSC.2009.5289382

Guan, X., Yang, B., Chen, C., Dai, W., & Wang, Y. (2016). A comprehensive overview of cyber-physical systems: From perspective of feedback system. *IEEE/CAA Journal of Automatica Sinica*, *3*(1), 1–14. https://doi.org/10.1109/JAS.2016.7373766

Guo, H., Yu, Y., & Skitmore, M. (2017). Visualization technology-based construction safety management: A review. *Automation in Construction*, *73*, 135–144. https://doi.org/10.1016/j.autcon.2016.10.004

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges. *Perspectives on Psychological Science*, *11*(6), 838–854. https://doi.org/10.1177/1745691616650285

Hasan, M. M., Watling, C. N., & Larue, G. S. (2022). Physiological signal-based drowsiness detection using machine learning: Singular and hybrid signal approaches. *Journal of Safety Research*, *80*, 215–225. https://doi.org/10.1016/J.JSR.2021.12.001

Hasanzadeh, S., Esmaeili, B., & Dodd, M. D. (2016). Measuring Construction Workers' Real-Time Situation Awareness Using Mobile Eye-Tracking. *Construction Research Congress 2016: Old and New Construction Technologies Converge in Historic San Juan - Proceedings of the 2016 Construction Research Congress, CRC 2016*, 2894–2904. https://doi.org/10.1061/9780784479827.288

Hasanzadeh, S., Esmaeili, B., & Dodd, M. D. (2017). *Measuring the Impacts of Safety Knowledge on Construction Workers' Attentional Allocation and Hazard Detection Using Remote Eye-Tracking Technology Attribute-Based Safety Risk Assessment View project Z-Box Illusion View project*. https://doi.org/10.1061/(ASCE)

Hassan, R., Cohanim, B., de Weck, O., & Venter, G. (2005). A comparison of particle swarm optimization and the genetic algorithm. *Collection of Technical Papers - AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, *2*, 1138–1150. https://doi.org/10.2514/6.2005-1897

Hasty.ai. (2022). *Patience*. https://hasty.ai/docs/mp-wiki/training-parameters/patience

Hegazy, T., Abdel-Monem, M., & Saad, D. A. (2014). Framework for enhanced progress tracking and control of linear projects. *Engineering, Construction and Architectural Management*, *21*(1), 94–110. https://doi.org/10.1108/ECAM-08-2012-0080

HEINRICH, H. W. (1950). *Industrial Accident Prevention ... Third Edition*. McGraw-Hill.

Heo, S., Kwon, S., & Lee, J. (2021). Stress Detection with Single PPG Sensor by Orchestrating Multiple Denoising and Peak-Detecting Methods. *IEEE Access*, *9*, 47777–47785. https://doi.org/10.1109/ACCESS.2021.3060441

Hjortskov, N., Rissén, D., Blangsted, A. K., Fallentin, N., Lundberg, U., & Søgaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. *European Journal of Applied Physiology*, *92*(1–2), 84–89. https://doi.org/10.1007/S00421-004-1055-Z/TABLES/2

Hou, X., Liu, Y., Sourina, O., Tan, Y. R. E., Wang, L., & Mueller-Wittig, W. (2016). EEG Based Stress Monitoring. *Proceedings - 2015 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2015*, 3110–3115. https://doi.org/10.1109/SMC.2015.540

Jebelli, H., Choi, B., & Lee, S. H. (2019). Application of Wearable Biosensors to Construction Sites. I: Assessing Workers' Stress. *Journal of Construction Engineering and Management*, *145*(12). https://doi.org/10.1061/(ASCE)CO.1943-7862.0001729

Jebelli, H., Hwang, S., & Lee, S. H. (2018). EEG-based workers' stress recognition at construction sites. *Automation in Construction*, *93*, 315–324. https://doi.org/10.1016/j.autcon.2018.05.027

Jyh-Shing Roger Jang, C.-T. S. E. M. (1997). *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Prentice Hall.

https://books.google.ca/books/about/Neuro_fuzzy_and_Soft_Computing.html?id=vN5QAAAAMAAJ&source=kp_book_description&redir_esc=y

Kamel, T. S., Hassan, M. A. M., & Morshedy, A. E.-. (2009). Advanced distance protection scheme for long Transmission lines in Electric Power systems using multiple classified ANFIS networks. *2009 Fifth International Conference on Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control*, 1–5. https://doi.org/10.1109/ICSCCW.2009.5379442

Karthikeyan, P., Murugappan, M., & Yaacob, S. (2011). A review on stress inducement stimuli for assessing human stress using physiological signals. *Proceedings - 2011 IEEE 7th International Colloquium on Signal Processing and Its Applications, CSPA 2011*, 420–425. https://doi.org/10.1109/CSPA.2011.5759914

Kerr, B., Condon, S. M., & McDonald, L. A. (1985). Cognitive Spatial Processing and the Regulation of Posture. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(5), 617–622. https://doi.org/10.1037/0096-1523.11.5.617

Khosravi, Y., Asilian-Mahabadi, H., Hajizadeh, E., Hassanzadeh-Rangi, N., Bastani, H., & Behzadan, A. H. (2015). Factors Influencing Unsafe Behaviors and Accidents on Construction Sites: A Review. *Https://Doi.Org/10.1080/10803548.2014.11077023*, *20*(1), 111–125. https://doi.org/10.1080/10803548.2014.11077023

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. https://arxiv.org/abs/1412.6980v9

Kothgassner, O. D., Felnhofer, A., Hlavacs, H., Beutl, L., Palme, R., Kryspin-Exner, I., & Glenk, L. M. (2016). Salivary cortisol and cardiovascular reactivity to a public speaking task in a virtual and real-life environment. *Computers in Human Behavior*, *62*, 124–135. https://doi.org/10.1016/j.chb.2016.03.081

Lackner, H. K., Papousek, I., Batzel, J. J., Roessler, A., Scharfetter, H., & Hinghofer-Szalkay, H. (2011). Phase synchronization of hemodynamic variables and respiration during mental challenge. *International Journal of Psychophysiology*, *79*(3), 401–409. https://doi.org/10.1016/J.IJPSYCHO.2011.01.001

Langdon, R. R., & Sawang, S. (2018). Construction Workers' Well-Being: What Leads to Depression, Anxiety, and Stress? *Journal of Construction Engineering and Management*, *144*(2), 1–15. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001406

Lean, Y., & Shan, F. (2012). Brief review on physiological and biochemical evaluations of human mental workload. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *22*(3), 177–187. https://doi.org/10.1002/hfm.20269

Lee, W., Lin, K. Y., Seto, E., & Migliaccio, G. C. (2017). Wearable sensors for monitoring on-duty and off-duty worker physiological status and activities in construction. *Automation in Construction*, *83*(June), 341–353. https://doi.org/10.1016/j.autcon.2017.06.012

Lei, Y., He, Z., Zi, Y., & Hu, Q. (2007). Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs. *Elsevier, Mechanical Systems and Signal Processing* . https://www.sciencedirect.com/science/article/pii/S0888327006002512

Leitão, P., Colombo, A. W., & Karnouskos, S. (2016). Industrial automation based on cyber-physical systems technologies: Prototype implementations and challenges. *Computers in Industry*, *81*, 11–25. https://doi.org/10.1016/j.compind.2015.08.004

Leung, M. Y., Chan, Y. S., & Yuen, K. W. (2010). Impacts of stressors and stress on the injury incidents of construction workers in Hong Kong. *Journal of Construction Engineering and Management*, *136*(10), 1093–1103. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000216

Leung, M. Y., Liang, Q., & Olomolaiye, P. (2016). Impact of Job Stressors and Stress on the Safety Behavior and Accidents of Construction Workers. *Journal of Management in Engineering*, *32*(1), 1–10. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000373

Lew, E., Chavarriaga, R., Silvoni, S., & Millán, J. del R. (2012). Detection of self-paced reaching movement intention from EEG signals. *Frontiers in Neuroengineering*, *JULY*. https://doi.org/10.3389/fneng.2012.00013

Li, J., Yang, Y., & Fu, G. (2011). Camera self-calibration method based on GA-PSO algorithm. *CCIS2011 - Proceedings: 2011 IEEE International Conference on Cloud Computing and Intelligence Systems*, 149–152. https://doi.org/10.1109/CCIS.2011.6045050

Lin, L., & Su, J. (2019). Anomaly detection method for sensor network data streams based on sliding window sampling and optimized clustering. *Safety Science*, *118*, 70–75. https://doi.org/10.1016/J.SSCI.2019.04.047

Liu, K., & Golparvar-Fard, M. (2015). Crowdsourcing Construction Activity Analysis from Jobsite Video Streams. *Journal of Construction Engineering and Management*, *141*(11), 04015035. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001010

Liu, Z., Yang, D. S., Wen, D., Zhang, W. M., & Mao, W. (2011). Cyber-physical-social systems for command and control. *IEEE Intelligent Systems*, *26*(4), 92–96. https://doi.org/10.1109/MIS.2011.69

Love, P. E. D., Edwards, D. J., & Irani, Z. (2010). Work Stress, Support, and Mental Health in Construction. *Journal of Construction Engineering and Management*, *136*(6), 650–658. https://doi.org/10.1061/(asce)co.1943-7862.0000165

Luo, Y., Xiao, Y., Cheng, L., Peng, G., & Yao, D. D. (2021). Deep Learning-based Anomaly Detection in Cyber-physical Systems. *ACM Computing Surveys (CSUR)*, *54*(5), 106. https://doi.org/10.1145/3453155

Mamdani, E., & Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies* , *7*(1), 1–13. https://www.sciencedirect.com/science/article/pii/S0020737375800022

Manchi, G. B., Gowda, S., & Hanspal, J. S. (2013). Study on Cognitive Approach to Human Error and its Application to Reduce the Accidents at Workplace. In *International Journal of Engineering and Advanced Texhnology(IJEAT)* (Vol. 2, Issue 6, pp. 236–242).

Maurice Clerc. (2010). *Particle swarm optimization*. John Wiley & Sons, 2010.

Munir, M., Siddiqui, S. A., Dengel, A., & Ahmed, S. (2019). DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series. *IEEE Access*, *7*, 1991–2005. https://doi.org/10.1109/ACCESS.2018.2886457

Munir, S., Stankovic, J. a., Liang, C.-J. M., & Lin, S. (2013). Cyber Physical System Challenges for Human-in-the-Loop Control. *The 8th International Workshop on Feedback Computing*.

Nalluri, M. S. R., Kannan, K., Manisha, M., & Roy, D. S. (2017). Hybrid Disease Diagnosis Using Multiobjective Optimization with Evolutionary Parameter Optimization. *Journal of Healthcare Engineering*. https://doi.org/10.1155/2017/5907264

Nath, R. K., Thapliyal, H., Caban-Holt, A., & Mohanty, S. P. (2020). Machine Learning Based Solutions for Real-Time Stress Monitoring. *IEEE Consumer Electronics Magazine*, *9*(5), 34–41. https://doi.org/10.1109/MCE.2020.2993427

Nunes, D., Silva, J. S., & Boavida, F. (2018). *A Practical Introduction to Human-in-the-Loop Cyber-Physical Systems*. John Wiley & Sons, Ltd.

Pakarinen, T., Pietila, J., & Nieminen, H. (2019). Prediction of Self-Perceived Stress and Arousal Based on Electrodermal Activity∗. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2191–2195. https://doi.org/10.1109/EMBC.2019.8857621

Palupi, D., Siti, R. ·, Shamsuddin, M., Siti, ·, Yuhaniz, S., Rini, D. P., Shamsuddin, · S M, Yuhaniz, · S S, Shamsuddin, S. M., & Yuhaniz, S. S. (2016). *Particle swarm optimization for ANFIS interpretability and accuracy*. *20*, 251–262. https://doi.org/10.1007/s00500-014-1498-z

Piccolo, D. (1990). A DISTANCE MEASURE FOR CLASSIFYING ARIMA MODELS. *Journal of Time Series Analysis*, *11*(2), 153–164. https://doi.org/10.1111/j.1467-9892.1990.tb00048.x

Ponce, P., del Arco, A., & Loprinzi, P. (2019). Physical Activity versus Psychological Stress: Effects on Salivary Cortisol and Working Memory Performance. *Medicina 2019, Vol. 55, Page 119*, *55*(5), 119. https://doi.org/10.3390/MEDICINA55050119

Powell, R., & Copping, A. (2010). Sleep deprivation and its consequences in construction workers. *Journal of Construction Engineering and Management*, *136*(10), 1086–1092. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000211

Přibil, J., Přibilová, A., & Frollo, I. (2020). First-Step PPG Signal Analysis for Evaluation of Stress Induced during Scanning in the Open-Air MRI Device. *Sensors 2020, Vol. 20, Page 3532*, *20*(12), 3532. https://doi.org/10.3390/S20123532

Qasem, S. N., & Shamsuddin, S. M. (2009). Generalization Improvement of Radial Basis Function Network Based on Multi-Objective Particle Swarm Optimization. *Journal of Artificial Intelligence*, *3*(1), 1–16. https://doi.org/10.3923/jai.2010.1.16

Rani, P., Sims, J., Brackin, R., & Sarkar, N. (2002). Online stress detection using psychophysiological signals for implicit human-robot cooperation. *Robotica*, *20*(6), 673–685. https://doi.org/10.1017/S0263574702004484

Rao, S. M., Losinski, G., Mourany, L., Schindler, D., Mamone, B., Reece, C., Kemeny, D., Narayanan, S., Miller, D. M., Bethoux, F., Bermel, R. A., Rudick, R., & Alberts, J. (2017). Processing speed test: Validation of a self-administered, iPad®-based tool for screening cognitive dysfunction in a clinic setting: *Https://Doi.Org/10.1177/1352458516688955*, *23*(14), 1929–1937. https://doi.org/10.1177/1352458516688955

Razaque, A., Abenova, M., Alotaibi, M., Alotaibi, B., Alshammari, H., Hariri, S., & Alotaibi, A. (2022). Anomaly Detection Paradigm for Multivariate Time Series Data Mining for Healthcare. *Applied Sciences 2022, Vol. 12, Page 8902*, *12*(17), 8902. https://doi.org/10.3390/APP12178902

Reason, J. (1991). Human error. In *Choice Reviews Online* (Vol. 29, Issue 02). Cambridge university press. https://doi.org/10.5860/choice.29-1194

Reiss, S., Peterson, R. A., Gursky, D. M., & McNally, R. J. (1986). Anxiety sensitivity, anxiety frequency and the prediction of fearfulness. *Behaviour Research and Therapy*, *24*(1), 1–8. https://doi.org/10.1016/0005-7967(86)90143-9

Renaud, P., & Blondin, J. P. (1997). The stress of Stroop performance: physiological and emotional responses to color–word interference, task pacing, and pacing speed. *International Journal of Psychophysiology*, *27*(2), 87–97. https://doi.org/10.1016/S0167-8760(97)00049-4

Renaud, P., Psychophysiology, J. B.-I. J. of, & 1997, undefined. (n.d.). The stress of Stroop performance: Physiological and emotional responses to color–word interference, task pacing, and pacing speed. *Elsevier*. Retrieved September 14, 2021, from https://www.sciencedirect.com/science/article/pii/S0167876097000494

Robati, F. N., & Iranmanesh, S. (2020). Inflation rate modeling: Adaptive neuro-fuzzy inference system approach and particle swarm optimization algorithm (ANFIS-PSO). *MethodsX*, *7*. https://doi.org/10.1016/J.MEX.2020.101062

Rodríguez-Arce, J., Lara-Flores, L., Portillo-Rodríguez, O., & Martínez-Méndez, R. (2020). Towards an anxiety and stress recognition system for academic environments based on physiological features. *Computer Methods and Programs in Biomedicine*, *190*. https://doi.org/10.1016/j.cmpb.2020.105408

Ruikar, K., Kotecha, K., Sandbhor, S., Thomas, A., Akanmu, A. A., Anumba, C. J., & Ogunseiju, O. O. (2021). Special issue: 'Next Generation ICT-How distant is ubiquitous computing?

Journal of Information Technology in Construction (ITcon), *26*, 506. https://doi.org/10.36680/j.itcon.2021.027

Saeed, S. M. U., Anwar, S. M., Khalid, H., Majid, M., & Bagci, U. (2020). EEG based classification of long-term stress using psychological labeling. *Sensors (Switzerland)*, *20*(7). https://doi.org/10.3390/s20071886

Salahuddin, L., Cho, J., Jeong, M. G., & Kim, D. (2007). Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, 4656–4659. https://doi.org/10.1109/IEMBS.2007.4353378

Schirner, G., Erdogmus, D., Chowdhury, K., & Padir, T. (2013). The future of human-in-the-loop cyber-physical systems. *Computer*, *46*(1), 36–45. https://doi.org/10.1109/MC.2013.31

Schneider, Patrick., & Xhafa, Fatos. (2022). *Anomaly detection and complex event processing over IoT data streams with application to eHealth and patient data monitoring*. Academic Press. http://www.sciencedirect.com:5070/book/9780128238189/anomaly-detection-and-complex-event-processing-over-iot-data-streams

Sevil, M., Rashid, M., Hajizadeh, I., Askari, M. R., Hobbs, N., Brandt, R., Park, M., Quinn, L., & Cinar, A. (2021). Discrimination of simultaneous psychological and physical stressors using wristband biosignals. *Computer Methods and Programs in Biomedicine*, *199*, 105898. https://doi.org/10.1016/j.cmpb.2020.105898

Sevil, M., Rashid, M., Hajizadeh, I., Park, M., Quinn, L., & Cinar, A. (2021). Physical activity and psychological stress detection and assessment of their effects on glucose concentration

predictions in diabetes management. *IEEE Transactions on Biomedical Engineering*, *68*(7), 2251–2260. https://doi.org/10.1109/TBME.2020.3049109

Shakerian, S., Habibnezhad, M., Ojha, A., Lee, G., Liu, Y., Jebelli, H., & Lee, S. H. (2021). Assessing occupational risk of heat stress at construction: A worker-centric wearable sensor-based approach. *Safety Science*, *142*, 105395. https://doi.org/10.1016/J.SSCI.2021.105395

Subhani, A. R., Mumtaz, W., Saad, M. N. B. M., Kamel, N., & Malik, A. S. (2017a). Machine learning framework for the detection of mental stress at multiple levels. *IEEE Access*, *5*, 13545–13556. https://doi.org/10.1109/ACCESS.2017.2723622

Subhani, A. R., Mumtaz, W., Saad, M. N. B. M., Kamel, N., & Malik, A. S. (2017b). Machine learning framework for the detection of mental stress at multiple levels. *IEEE Access*, *5*, 13545–13556. https://doi.org/10.1109/ACCESS.2017.2723622

Subhani, A. R., Mumtaz, W., Saad, M. N. B. M., Kamel, N., & Malik, A. S. (2017c). Machine learning framework for the detection of mental stress at multiple levels. *IEEE Access*, *5*, 13545–13556. https://doi.org/10.1109/ACCESS.2017.2723622

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *EEE Transactions on Systems, Man, and Cybernetics*, *1*, 116–132. https://ieeexplore.ieee.org/abstract/document/6313399/

Teng, M. (2010). Anomaly detection on time series. *Proceedings of the 2010 IEEE International Conference on Progress in Informatics and Computing, PIC 2010*, *1*, 603–608. https://doi.org/10.1109/PIC.2010.5687485

Teplan, M. (2002). Fundamentals of EEG measurement. *Measurement Science Review*, *2*(2), 1–11.

Terry, J. K., Jayakumar, M., & de Alwis, K. (2021). *Statistically Significant Stopping of Neural Network Training*. https://doi.org/10.48550/arxiv.2103.01205

Tibshirani, R., & Wang, P. (2008). Spatial smoothing and hot spot detection for CGH data using the fused lasso. *Biostatistics*, *9*(1), 18–29. https://doi.org/10.1093/BIOSTATISTICS/KXM013

Toole, T. M. (2002). Construction Site Safety Roles أدوار سلامة موقع البناء. *Journal of Construction Engineering and Management*, *128*(3), 203–210. https://doi.org/10.1061/(ASCE)0733-9364(2002)128

Tulen, J. H. M., Moleman, P., van Steenis, H. G., & Boomsma, F. (1989). Characterization of stress reactions to the Stroop Color Word Test. *Pharmacology Biochemistry and Behavior*, *32*(1), 9–15. https://doi.org/10.1016/0091-3057(89)90204-9

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

Vrijkotte, T. G. M., Van Doornen, L. J. P., & De Geus, E. J. C. (2000). Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension*, *35*(4), 880–886. https://doi.org/10.1161/01.HYP.35.4.880

Wang, D., Chen, J., Zhao, D., Dai, F., Zheng, C., & Wu, X. (2017). Monitoring workers' attention and vigilance in construction activities through a wireless and wearable electroencephalography system. *Automation in Construction*, *82*, 122–137. https://doi.org/10.1016/j.autcon.2017.02.001

Wang, J., Fang, K., Pang, W., Electrical, J. S.-J. of, & 2017, undefined. (2017). Wind power interval prediction based on improved PSO and BP neural network. *Koreascience.or.Kr*, *12*(3), 989–995. https://doi.org/10.5370/JEET.2017.12.3.989

Wang, J., & Razavi, S. (2019). Integrated and Automated Systems for Safe Construction Sites. *Professional Safety*, *64*(02), 41–45. https://doi.org/ASSE-19-02-41

Wittchen, H. U., & Boyer, P. (1998). Screening for anxiety disorders. Sensitivity and specificity of the Anxiety Screening Questionnaire (ASQ-15). *British Journal of Psychiatry*, *173*(JULY SUPPL. 34), 10–17. https://doi.org/10.1192/s000712500029346x

Wood, A., & Stankovic JA. (2008). Human in the loop: Distributed data streams for immersive cyber-physical systems. *ACM SIGBED*. https://dl.acm.org/doi/abs/10.1145/1366283.1366303?casa_token=4APQVywBAQwAAAAA:_FcvtX7q3JwK9PHBQF4BcZlZ0983FDHDbI1lbJ8SLAwyjxkdsm7GfMhJQ2i5QYGlZ8hzWvKZ17y1mg

Xie, J., Liu, S., & Wang, X. (2022). Framework for a closed-loop cooperative human Cyber-Physical System for the mining industry driven by VR and AR: MHCPS. *Computers & Industrial Engineering*, *168*, 108050. https://doi.org/10.1016/J.CIE.2022.108050

Yadav, M., Sakib, M. N., Nirjhar, E. H., Feng, K., Behzadan, A., & Chaspari, T. (2020). Exploring individual differences of public speaking anxiety in real-life and virtual presentations. *IEEE Transactions on Affective Computing*. https://doi.org/10.1109/TAFFC.2020.3048299

Yu, J., Wang, S., & Xi, L. (2008). Evolving artificial neural networks using an improved PSO and DPSO. *Neurocomputing*, *71*(4–6), 1054–1060. https://doi.org/10.1016/J.NEUCOM.2007.10.013

Yuqi, L. (2021). *Colab Pro vs. Free — AI Computing Performance*. Towards Data Science. https://towardsdatascience.com/colab-pro-vs-free-ai-computing-performance-4e983d578fb2

Zhang, C., Song, D., Chen, Y., Feng, X., Lumezanu, C., Cheng, W., Ni, J., Zong, B., Chen, H., & Chawla, N. v. (2019). A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data. *33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019*, *33*(01), 1409–1416. https://doi.org/10.1609/aaai.v33i01.33011409

Zhang, C., Tang, P., Cooke, N., Buchanan, V., Yilmaz, A., st. Germain, S. W., Boring, R. L., Akca-Hobbins, S., & Gupta, A. (2017). Human-centered automation for resilient nuclear power plant outage control. *Automation in Construction*, *82*(March), 179–192. https://doi.org/10.1016/j.autcon.2017.05.001

Zhang, Y., & Wu, L. (2011). A hybrid TS-PSO optimization algorithm. *Journal of Convergence Information Technology*,. https://www.academia.edu/download/34038554/A_Hybrid_TS-PSO_Optimization_Algorithm.pdf

Zokaei, M., Jafari, M. J., Khosrowabadi, R., Nahvi, A., Khodakarim, S., & Pouyakian, M. (2020). Tracing the physiological response and behavioral performance of drivers at different levels of mental workload using driving simulators. *Journal of Safety Research*, *72*, 213–223. https://doi.org/10.1016/J.JSR.2019.12.022